THE PROCEEDINGS OF
THE PHYSICAL SOCIETY

VOL. 47, PART 1 January 1, 1935 No. 258

621.396.615.1 621.385.16

ELECTRONIC THEORY AND THE MAGNETRON OSCILLATOR

By W. E. BENHAM

Received August 9, 1934. Read November 2, 1934.

ABSTRACT. The present paper extends previous analysis* to cover any degree of space-charge limitation. In addition to terms involving initial velocities and accelerations of electrons, the effect of a magnetic field of constant value has been included.

The potential-distribution is shown to be but slightly influenced by the magnetic field under steady-state conditions. It is shown that Langmuir's results for this case† still apply fairly accurately for potential-minimum calculations in the presence of a magnetic field. Particular attention is given to the value of the electric field at the cathode and to the average initial velocity of those electrons which reach the potential-minimum and thus constitute the anode current. The prolongation of electron transit time by the magnetic field is fully discussed.

The electron-velocity is composed of a forward component parallel to the electric field and a drift component at right angles to both electric and magnetic fields. The forward velocity may be oscillatory in character but the drift velocity is unaffected by alternating fields.

The general solution for the forward velocity contains two arbitrary functions the correct choice of which is fully discussed. The inclusion of a magnetic field is definitely of assistance in establishing that the choice of the arbitrary functions is governed mainly by the necessity for avoiding infinite values, rather than by boundary conditions of the usual type. This anomalous state of affairs arises from the non-linearity of the original equations, which involves the introduction at an early stage of the arbitrarily assigned boundary values of the steady-state components of velocity and acceleration as well as those of the direct and alternating current.

Formulae are given for the components normal and parallel to the plates of the electron-velocity averaged over all electrons, and application is made to a real and to a virtual cathode.

§ 1. INTRODUCTION

In the vacuum-tube theory of some twenty years ago it was considered sufficient to know the relations existing between direct currents and voltages, as obtained theoretically or by direct measurement, to be well acquainted with the processes that take place in high-vacuum tubes. At the present day the same outlook is still

* W. E. Benham, *Phil. Mag.* **11**, 457 (1931). † *Phys. Rev.* **21**, 419 (1923).

very general, and not without abundant experimental justification. In the design of vacuum tubes for broadcast or commercial transmission and reception a number of constants have now, of course, to be determined in addition to the static characteristic curves. The most important of these are the electrostatic capacities between the several pairs of electrodes associated with the tube. Where curves other than the static characteristic curves are quoted, such as those showing audio output power and efficiency as a function of output impedance, such curves can generally be regarded as known from the static characteristics, though in practice dynamic measurements are usual where the saving of time is a major consideration.

It is generally true that the higher the frequency of the alternating current for which a given vacuum tube is designed the greater is the uncertainty that attaches to any estimate of the performance based on static characteristics. In addition to the limitations of efficiency associated with interelectrode capacities already referred to, dielectric and eddy-current losses may be paramount. In many cases the electron-inertia is rightly considered negligible in comparison with other factors. What is not generally appreciated, however, is that electron-inertia may be of importance even at audio frequencies. For example, the operating interelectrode capacities are known to depend on the space-charge conditions between the electrodes. If the electrons were possessed of charge but were fixed in space, they would invariably increase the capacity as compared with the electrostatic value. Owing to the finite mass of the electron the capacity is lessened by an amount depending on the space-current conditions and geometry of the electrode-pair considered, and only to a minor extent on the frequency (except at frequencies
T comparable with T^{-1}, where T is transit time of electrons).

It is, admittedly, difficult to predict with certainty the effect of space-charge inertia in multi-electrode valves. It is possible, however, to say with confidence that a stage is rapidly approaching where greater attention will have to be paid to this question. In the pentagrid tube, for example, we have a virtual cathode formed between a pair of grids and serving as electron-source to succeeding grids.

It appears that a virtual cathode is possessed of remarkable properties, accurate information concerning which can only be obtained by means of an exhaustive theoretical study covering electron-inertia, together with the initial velocities and accelerations at the real cathode.

In the attempt to establish some of the properties of a magnetron oscillator, a magnetic field applied parallel to the plates of a parallel plane diode was studied, first of all by neglecting initial velocities and accelerations, and anomalous results were obtained at the critical plane, which is the turning-point of electrons in the diode. Now the critical plane in a magnetron differs from a virtual cathode* only in respect of the existence of a drift-velocity component parallel to the plates. A study of the critical plane should therefore result in information which can be applied to virtual cathodes in general. With a view to obtaining further information, initial velocities and accelerations have been included, and the general solution has been derived.

* L. Tonks, *Phys. Rev.* **29**, 913 (1927).

§ 2. NATURE OF THE PROBLEM

(2.1) *General conceptions.* The problem involved in generalized electronic theory is very similar in nature to that of the motion of charged particles in free space under the influence of electromagnetic waves, but with one important difference. In illustration of this difference we may consider two equal electromagnetic waves travelling in opposite directions. Let them meet at a point. If the electric vectors of the two waves are parallel, a quasistationary condition, in which the magnetic forces in the waves cancel one another, will obtain at the point considered and also along the line of overlap of the waves. Further, if the waves considered have travelled sufficiently from their respective source to be considered plane, they will meet in a plane and a volume will be traced out in which the above quasistationary state applies.

If in this volume we insert a plane thermionic cathode in a condition to emit electrons, and orient the cathode at right angles to the electric vectors of the waves, the motions of the electrons in the space will be governed by the same considerations as apply in the practical case of a valve subjected to alternating potentials.

The illustration serves to show that the problem is somewhat simplified as compared with the general electromagnetic problem, which has hitherto defied solution, in that terms representing the magnetic vectors may be dropped out of the equations. This simplification is fortunate in that the complication is already considerable in view of the mutual repulsions of the electrons. Secondary magnetic effects due to the motions of the electrons themselves are small for velocities low compared with c, and these will be neglected along with the magnetic vectors, the only magnetic field considered being one of constant value.

Still another conception which it is desirable to introduce at an early stage is the subdivision into alternating and direct-current components of certain of the physical quantities appearing in the equations. This operation is effected in the ordinary course of the analysis, but a word as to its justification may not be out of place here. In previous work the vector quantities all occurred in the same direction. Any vector could therefore be subdivided into components of a Fourier series which could be added together at any point where desired. In the present case we have to justify the operation of choosing components of the physical quantities parallel to the axes of co-ordinates and afterwards effecting the harmonic subdivision. In this connection we have to deal with the non-linear expression $U\nabla . U$, which, when components are taken parallel to the axes, represents nine terms in all. While the velocity U of the electrons is itself a vector, $U\nabla$ denotes the scalar operator

$$\left(U_x\frac{\partial}{\partial x}+U_y\frac{\partial}{\partial y}+U_z\frac{\partial}{\partial z}\right).$$

The x-component of $U\nabla . U$ is thus

$$\left(U_x\frac{\partial}{\partial x}+U_y\frac{\partial}{\partial y}+U_z\frac{\partial}{\partial z}\right)U_x.$$

If we effect harmonic subdivision before dividing into components parallel to the axes, we obtain ($\bar{U}$ being independent of t)

$$U\nabla . U = \bar{U}\nabla . \bar{U} + (\bar{U}\nabla . u + u\nabla . \bar{U}) + u\nabla . u.$$

The x-component of $(\bar{U}\nabla . u + u\nabla . \bar{U})$ is

$$\left(\bar{U}_x \frac{\partial}{\partial x} + \bar{U}_y \frac{\partial}{\partial y} + \bar{U}_z \frac{\partial}{\partial z}\right) u_x + \left(u_x \frac{\partial}{\partial x} + u_y \frac{\partial}{\partial y} + u_z \frac{\partial}{\partial z}\right) \bar{U}_x$$

$$= \frac{\partial}{\partial x}(\bar{U}_x u_x) + \left(\bar{U}_y \frac{\partial u_x}{\partial y} + u_y \frac{\partial \bar{U}_x}{\partial y}\right) + \left(\bar{U}_z \frac{\partial u_x}{\partial z} + u_z \frac{\partial \bar{U}_x}{\partial z}\right),$$

the result which is also obtained by choosing components first and afterwards effecting subdivision. We may therefore consider the method of harmonic subdivision as justified when used without further comment in § 5 at (5·1).

A few preliminary considerations will now suffice to pave the way for the analysis.

(2.2) *Notation.* In the case of the treatment applicable to *individual* electrons (2.4) the notation used is quite straightforward. The use of fluxional notation to indicate differentiation following the motion of the electron is quite usual and rather more satisfactory than the use of total differential notation d/dt, since the differentiation is, in fact, partial,

$$\dot{x} = \frac{\partial x}{\partial t}, \quad \dot{y} = \frac{\partial y}{\partial t}, \quad \dot{z} = \frac{\partial z}{\partial t},$$

$$\frac{\partial}{\partial t} \equiv \left.\frac{\partial}{\partial t}\right|_{t_0},$$

t_0 where t_0 (which $< t$) represents the instant at which the electron in question left the origin. Thus only in the time-steady state can $\dot{x}$, $\dot{y}$, $\dot{z}$ be regarded as total differentials.

In the *general* case, the coordinate system used is

$$(x, y, z, t),$$

, z where x, y, z are now independent variables, and $\partial/\partial t$ denotes partial differentiation with respect to time at constant (x, y, z). The velocity in the general case is repre-
U_z sented by U having components U_x, U_y, U_z parallel to the axes. At a later stage, when U_x comes in for special consideration, the suffix x is omitted. This does not
o lead to confusion and permits of the use of the suffix o to denote forward velocity of emission. The various components of U in the harmonic sense are represented
u' by subdivision of U into components of zero order $\bar{U}$, first order u, second order u' and so on.

(2.3) *No restriction to small amplitudes.* The values of the harmonic components
J of U are determined by assigning to the total current J the wave form

, p $$J = \bar{J} + j_1 \sin pt + j_2 \sin^2 pt + \ldots,$$

the number of available equations being equal to the number of harmonics it is desired to consider. Thus, at any selected value of x

f $$\bar{U} = f(\bar{J}),$$
g $$u = g(\bar{U}, \bar{J}, j, t),$$
h $$u' = h(\bar{U}, \bar{J}, u_1, j_1, j_2, t),$$

and so on. In general it is only necessary to determine the functions f and g. The function h was determined for planes and cylinders in the absence of a magnetic field, but the calculation is highly laborious. In cases where we are investigating the ability of a valve to give rise to oscillations it may safely be assumed that the tube and associated circuit will distinguish between harmonics and fundamental, so that we should always be able to pick out the latter. In most cases, therefore, determination of the properties of the fundamental component gives all the information necessary to the problem, and it is to be noted that with the inclusion of initial velocities and accelerations the restriction $u \ll U$ disappears, the solution obtained being entirely independent of this assumption except when applied at points sufficiently close to the anode, for the electron current to be cut off during part of the cycle. This only occurs for magnetic fields in the neighbourhood of the critical value.

The above applies equally to the solution for individual electrons, but in this case the solution is inadequate for other reasons; see (2.6). The treatment outlined in (2.4) is given chiefly in order to facilitate understanding of the main problem.

(2.4) *Equations and solution for individual electrons of the space charge.* Consider the motion of charged particles under the combined influence of a uniform magnetic field H and an electric field X which will be non-uniform in general. Let the axis of z be taken along H and that of x along X. If we choose a right-handed system (x, y, z), the equations of motion of any selected particle are

$$\left.\begin{aligned} m\ddot{x} &= eX + eH\dot{y} \\ m\ddot{y} &= -eH\dot{x} \\ m\ddot{z} &= 0 \end{aligned}\right\} \qquad \ldots\ldots(1),$$

where m and e are the mass and negative charge of an electron.

The four cases to be considered are as follows: (a) X uniform and constant; (b) X uniform, but varying in time; (c) X non-uniform, but independent of time at a given value of x; (d) X non-uniform and varying with time.

The first two cases are dealt with in (2.41), the second two in (2.42).

(2.41) *Space-charge forces negligible.* (a) *X uniform and constant.* The integrals of (1) are as follows:

$$\dot{x} = \dot{x}_0 \cos \omega t + \dot{y}_0 \sin \omega t + \frac{X}{H} \sin \omega t \qquad \ldots\ldots(2),$$

$$\dot{y} = -\dot{x}_0 \sin \omega t + \dot{y}_0 \cos \omega t - \frac{X}{H}(1 - \cos \omega t) \qquad \ldots\ldots(3),$$

$$\dot{z} = 0,$$

in which the suffix 0 corresponds to initial velocity components, and ω is written for $(e/m) H$. The quantity ω represents the natural angular frequency of rotation of electrons about the lines of magnetic force, and will require such frequent discussion that for the future we shall refer to ω as the *field frequency*, for want of a better term. In this and in other connections the word "frequency" will be freely used where "angular frequency" is to be understood.

A further integration gives the path of the particles:

$$x=\frac{\dot{x}_0}{\omega}\sin\omega t+\frac{1}{\omega}\left(\dot{y}_0+\frac{X}{H}\right)(1-\cos\omega t) \qquad \text{......(4)},$$

$$y=\frac{\dot{x}_0}{\omega}\cos\omega t-\frac{\dot{y}_0}{\omega}\sin\omega t-\frac{X}{\omega H}(\omega t-\sin\omega t) \qquad \text{......(5)},$$

$$z=\dot{z}_0 t.$$

We note from (4) the effect of tangentially emitted electrons on the forward path (that in the x direction). It can be shown that the maximum value of x is affected more by the $\dot{y}_0$ term than by the $\dot{x}_0$ term.

The x component of acceleration at the origin is given by

$$\ddot{x}_0=\omega\left(\dot{y}_0+\frac{X}{H}\right)=\omega\dot{y}_0+\frac{e}{m}X \qquad \text{......(6)}.$$

By equation (3) there is a mean transverse velocity of drift given by $-X/H$. This was referred to by Chapman as *the drift velocity of the free charge**. Chapman also drew attention to the fact that the mean value of $\dot{x}$ was zero despite the electric field being in the direction of x. This conclusion corresponds, of course, to perfect freedom of motion of the particles. In the case of a valve the anode may collect electrons so that the cycloidal path is interrupted, and in this case there will be a resultant motion of electrons in the x-direction.

By direct integration of the second equation of (1) we see that

$$\dot{y}=\dot{y}_0-\omega x \qquad \text{......(7)}.$$

This equation shows how the drift velocity varies with distance from the origin. It also shows that if we take up a position of observation at any selected value of x and watch charges moving in the yz plane, the average effect observed will be that of a constant drift velocity. It is to be noted that equation (7) holds good in all the cases (*a*) to (*d*) under consideration.

(*b*) *X uniform, but varying in time.* Let

$$X=\bar{X}-X_1\cos pt \qquad \text{......(8)},$$

in which p will be referred to as *the impressed frequency*. Thus, in the concrete case of the magnetron oscillator, p would depend on the natural frequency of the associated circuit. On the other hand, p is also to be thought of as the frequency at which the system will oscillate, if it oscillates at all. We thus adopt the method of impressing a disturbance of frequency p on the system and studying the reactions of the system to that disturbance. The forward-velocity component may be obtained by solving for $\dot{x}$ by means of the equation

$$\dddot{x}+\omega^2\dot{x}=\frac{e}{m}pX_1\sin pt \qquad \text{......(9)}.$$

The solution obtained is

$$\dot{x}=\dot{x}_0\cos\omega(t-t_0)+\frac{\ddot{x}_0}{\omega}\sin\omega(t-t_0)+\frac{e}{m}pX_1\frac{1}{p^2-\omega^2}$$
$$\times\left[-\sin pt+\sin pt_0\cos\omega(t-t_0)+\frac{p}{\omega}\cos pt_0\sin\omega(t-t_0)\right] \qquad \text{......(10)},$$

* *Proc. roy. Soc.* A, **122**, 378 (1929).

in which t_0 represents the instant at which the electron in question left the origin. t_0

It must be noted that $\ddot{x}_0$ is now a function of t_0 given by substituting (8) in (1) and putting $t=t_0$

$$\ddot{x}_0=\omega\dot{y}_0+\frac{e}{m}(\bar{X}-X_1\cos pt_0).$$

(2.42) *Space-charge forces not negligible.* (*c*) *X non-uniform, but independent of time at a given value of x.*

The non-uniformity of the field may be expressed by the Maxwellian equation

$$\dot{X}=4\pi c^2 J \qquad \text{......(11)},$$

where J is Maxwell's total current. Since J will here be independent of time, the effect of space-charge is to give rise to a field X such that J

$$X=4\pi c^2\bar{J}.t+X_0,$$ $\bar{J}$

in which t_0 is taken as zero at the origin to indicate the time steady state.

Instead of equation (8) we require the equation

$$\ddot{x}+\omega^2\dot{x}=4\pi\bar{J} \qquad \text{......(12)},$$

the steady-state solution of which is

$$\dot{x}=\dot{x}_0\cos\omega t+\frac{\ddot{x}}{\omega}\sin\omega t+4\pi\frac{e}{m}c^2\frac{\bar{J}}{\omega^2}(1-\cos\omega t) \qquad \text{......(13)},$$

in which $\ddot{x}_0$ is given by equation (6).

If we put $\bar{J}$ equal to 0 equation (13) reduces to equation (2), as would be expected, since if there is no space current the number of electrons will be insufficient to give rise to a non-uniform field. It is important to note that $\bar{J}$ represents the contribution of all electrons to the space current: equation (11) is fundamental and applies exactly whether or no a velocity-distribution exists.

(*d*) *X non-uniform and varying with time.* This is the general case. Writing

$$J=\bar{J}+j_1\sin pt$$

in equation (12), we obtain the following solution:

$$\dot{x}=\dot{x}_0\cos\omega(t-t_0)+\frac{\ddot{x}_0}{\omega}\sin\omega(t-t_0)+4\pi\frac{e}{m}c^2\frac{\bar{J}}{\omega^2}\{1-\cos\omega(t-t_0)\}$$

$$+\frac{4\pi\frac{e}{m}c^2j_1}{p^2-\omega^2}\left[-\sin pt+\sin pt_0\cos\omega(t-t_0)+\frac{p}{\omega}\cos pt_0\sin\omega(t-t_0)\right] \qquad \text{...(14)}.$$

The value of $\ddot{x}_0$ is related to that of the field at the origin. We have, by integration of (11) with respect to t over the interval $(t-t_0)$,

$$X-X_0=4\pi c^2\int_{t_0}^{t}Jdt=4\pi c^2\left[\bar{J}(t-t_0)+\frac{j_1}{p}(\cos pt_0-\cos pt)\right] \qquad \text{...(15)}.$$

The right-hand side vanishes when $t=t_0$, and we see that X_0 remains indeterminate. This, however, would be expected since it is not sufficient to know the space current, but we require also to know the total emission. If n_0 be the number of

electrons emitted with forward components of velocity lying between $\dot{x}_0$ and $\dot{x}_0 + d\dot{x}_0$, and if J_0 denote total emission plus displacement current, both reckoned per cm^2 of cathode surface,

$$J_0 = \int_0^\infty n_0 e \, d\dot{x}_0 + \frac{1}{4\pi c^2} \frac{\partial X_0}{\partial t_0}\bigg|_{\text{cathode}} \qquad \text{......(16).}$$

Now the Schottky* effect results in a slight variation of total emission with field-strength, but the alternating field itself may be considered insufficient to be effective in this respect. Again, the total emission is subject to random variations known as "Schroteffekt"†; these variations, which amount to a very small fraction of the total emission, could if desired be represented by regarding J_0 as modulated by the current to which they give rise. For present purposes we may neglect both Schottky and Schroteffekts. Then n_0 will be independent of t_0, i.e. the space charge *immediately* in front of the cathode is independent of the alternating field at the cathode, a consideration which also follows from the inability of the electrons to respond to periodic forces immediately on emerging from the cathode. Thus n_0 and $\dot{x}_0$ will be independent of t_0. We may then split equation (16) into two equations

$$\bar{J}_0 = \int_0^\infty n_0 e \, d\dot{x}_0$$

and

$$j_1 \sin pt_0 = \frac{1}{4\pi c^2} \frac{\partial X_0}{\partial t_0}\bigg|_{\text{cathode}} \qquad \text{......(17).}$$

On integrating equation (17) with respect to t_0, and subject to the condition that $X_0 = \bar{X}_0$ when $t_0 = 0$,

$$X_0 = \bar{X}_0 + \frac{4\pi c^2 j_1}{p}(1 - \cos pt_0) \qquad \text{......(18).}$$

The acceleration at the cathode is given by

$$\ddot{x}_0 = \omega \dot{y}_0 + 4\pi \frac{e}{m} c^2 j_1 (1 - \cos pt_0) \frac{1}{p} \qquad \text{......(19).}$$

The electric field at any point of the system is given by equation (15), which with the aid of equation (18) becomes

$$X = \bar{X}_0 + 4\pi c^2 \left[\bar{J}(t - t_0) + \frac{j_1}{p}(1 - \cos pt) \right] \qquad \text{......(20).}$$

It will be seen that the field frequency ω does not enter into the expression, a result that appears to conflict with the potential-distribution obtained in § 4, which does depend on ω. The apparent paradox is explained by the alteration of $(t - t_0)$. The electric field as given by equation (20) is thus to be regarded as modified by a magnetic field through modification of the transit time.

(2.5) *Examination of the solution in particular cases.* When the field-frequency ω becomes equal to the impressed frequency p, casual inspection of the solutions (10) and (14) suggests a resonance phenomenon similar to that corresponding to dispersion phenomena in general. With due care in the taking of limits, however,

* Schottky, *Ann. Phys.*, Lpz., **44**, 1011 (1914).
† *Ibid.* **57**, 541 (1918); **68**, 157 (1922).

equation (14) is seen to remain finite in the case where $\omega = p$, and to assume the value

$$\dot{x} = \dot{x}_0 \cos p\,(t-t_0) + \frac{\ddot{x}_0}{p} \sin p\,(t-t_0) + 4\pi \frac{e}{m} c^2 \bar{J}\, \frac{1}{p^2}\{1-\cos p\,(t-t_0)\}$$

$$+ \frac{2\pi \frac{e}{m} c^2 \dot{j}_1}{p^2} [\cos pt_0 \sin \omega\,(t-t_0) - p\,(t-t_0) \cos pt] \quad \text{......(21).}$$

It will appear later that a form of resonance can occur when $\omega = p$. For the time being we note that equation (21) shows no signs of a marked resonance effect.

Another case of interest is that of small magnetic field, $\omega \ll p$. In this case equation (14) reduces to

$$\dot{x} = \dot{x}_0 + \ddot{x}_0\,(t-t_0) + 2\pi \frac{e}{m} c^2 \bar{J}\,(t-t_0)$$

$$+ \frac{4\pi \frac{e}{m} c^2 \dot{j}_1}{p^2} [-\sin pt + \sin pt_0 + p\,(t-t_0) \cos pt_0] \quad \text{......(22).}$$

The cases $\omega = p$, $\omega \ll p$ may be compared by considering the square-bracket terms of equations (21) and (22) in the particular case $t_0 = 0$. Writing

$$\beta = 4\pi \frac{e}{m} c^2 \dot{j}_1,$$ β

we have to compare the expressions

$$\frac{\beta}{2p^2} (\sin pt - pt \cos pt)$$

and

$$\frac{\beta}{p^2} (pt - \sin pt),$$

in which t is now to be regarded as the instantaneous value of the transit time. For small values of pt we have in each case t

$$\tfrac{1}{6}\beta\, pt^3,$$

i.e. the expressions are identical near the cathode. This result suggests that the magnetic field is without effect on the velocity of electrons near the cathode. This aspect of the solution is dealt with in § 4 at (4.5). For large values of t we obtain, if $pt \gg 1$,

$$\frac{\beta}{2p^2} (-pt \cos pt)$$

and

$$\frac{\beta}{p^2} (pt),$$

a result which shows that the amplitude of motion in the case of a small magnetic field may be double that obtaining in the case of a field-frequency equal to the impressed frequency.

(2.6) *The inadequacy of particle dynamics.* From the mathematical point of view equation (14) represents the solution of the whole problem, for in order to embrace all the electrons it would merely be necessary to sum the solution over all values of the emission-velocity components $\dot{x}_0$ and $\dot{y}_0$. Furthermore, if it is desired to express the solution in terms of (x, y, z, t) as independent variables, there is a technique of transformation for this outlined in appendix 1. Essentially, then, the problem is solved. When, however, we attempt to apply the technique of appendix 1

to the case under consideration, we find that it breaks down on grounds of intractability. It is therefore impracticable to convert from one coordinate system to the other except in special cases. It follows that if a solution is required in terms of (x, y, z, t) as independent variables, then it must be obtained by working throughout in these variables. It will be seen that the only disadvantage of working in terms of the variables (x, y, z, t), or equivalent variables, disappears when once a method has been found for handling the differential equations.

The inadequacy of the solution (14) as it stands arises from the existence of a variation time arising from fluctuations in the transit time $(t-t_0)$. What is required for vacuum-tube purposes is a solution which has a readily ascertainable value at a given value of (x, y, z).

§ 3. A GENERAL THEORY OF SPACE-CHARGE PHENOMENA IN A HIGH VACUUM

(3.1) *General equations for electron-motion on classical electromagnetic theory.* The medium in which the electrons are moving possesses unit permeability and dielectric constant prior to the introduction of electrons. Because of the presence of electrons some modification in the dielectric constant will result. On the concepts of the electron theory, however, such modification leaves the dielectric constant of the space between the electrons unaltered, so that in the equations which follow we shall put $\kappa=\mu=1$ everywhere. The effective change in dielectric constant comes about indirectly through the non-uniformity of the electric field resulting from the distribution of electrons, and since the effect of this non-uniformity is included in the first of the equations below it would be definitely incorrect to include a quantity κ to represent the altered dielectric constant, for then the effect would be included twice over.

In the six equations which follow all quantities are expressed in electromagnetic units.

V The equation satisfied by the electric potential V is

$$\nabla^2 V - \frac{1}{c^2}\frac{\partial^2 V}{\partial t^2} = -4\pi c^2 \rho \qquad \text{......(23)},$$

ρ where ρ is the density of distribution of electrons at point (x, y, z).

Maxwell's total current density J is given by

$$J = \rho U + \frac{1}{4\pi c^2}\frac{\partial E}{\partial t} \qquad \text{......(24)},$$

U, E where U is the velocity and E the electric field.

The equation of motion is

$$\frac{\partial U}{\partial t} + U\nabla . U = \frac{e}{m}\{E - [U, H]\} \qquad \text{......(25)},$$

H where H is the magnetic field.

We have also the two Maxwellian equations

$$4\pi J = \text{curl } H \qquad \text{......(26)},$$

$$-\frac{\partial H}{\partial t} = \text{curl } E \qquad \text{......(27)},$$

while the equation of continuity takes the simple form

$$\operatorname{div} J = 0 \qquad \text{......(28).}$$

Equations (23) to (28) are necessary equations for the determination of the six quantities V, ρ, J, U, E and H. It does not follow that six equations are mathematically sufficient, in view of the existence of partial differentials.

In § 2 at (2.1) it was pointed out that if the magnetic vectors of two electromagnetic waves meeting in space cancel one another the general problem would be considerably simplified. Now in equations (25) and (26) the magnetic vectors are included in H, and if we drop the magnetic vectors H will reduce to a known quantity, that representing the joint effect of the magnetic field set up by the electron current and that corresponding to the applied magnetic field. In order, however, to appreciate to the full the extent of the simplification consider the equation

$$E = -\frac{\partial \Omega}{\partial t} - \operatorname{grad} V \qquad \text{......(29),}$$

where Ω represents the magnetic vector potential.

Ω

If now Ω is equated to zero, equation (29) becomes

$$E = -\operatorname{grad} V \qquad \text{......(30).}$$

We may then eliminate E from equations (24), (25) and (27), thus reducing the number of unknowns to five, while still retaining six equations intact.

(3.2) *Case in which magnetic field is constant.* If the applied magnetic field be invariant in time, the only alternating magnetic field present will be that corresponding to secondary magnetic effects, and these are negligible for an electron-velocity low compared with c. We may then equate $\partial H/\partial t$ to zero in equation (27), obtaining with equation (30) the known result for scalar quantities,

$$\operatorname{curl}(\operatorname{grad} V) = 0.$$

Secondary electric effects are negligible for the same reason, which means that the electric potential in equation (23) need not be considered as a retarded potential, the term $c^{-2}\,\partial^2 V/\partial t^2$ may be omitted, and we are left with Poisson's equation. The last term of equation (24) also includes secondary electric effects, but is itself by no means negligible in the case where the displacement current between a pair of electrodes is contemplated. We shall not require equation (26), since the effect on the electrons themselves of the magnetic field produced by their own current is a secondary magnetic effect and as such is being considered negligible. Our general equations are now

$$\nabla^2 V = -4\pi c^2 . \rho \qquad \text{......(31),}$$

$$J = \rho U - \frac{1}{4\pi c^2}\frac{\partial}{\partial t}(\operatorname{grad} V) \qquad \text{......(32),}$$

$$\frac{\partial U}{\partial t} + U\nabla . U = -\frac{e}{m}\{\operatorname{grad} V + [U, H]\} \qquad \text{......(33),}$$

$$\operatorname{div} J = 0 \qquad \text{......(34).}$$

In our application of the above equations we shall restrict ourselves to the cases where H is known everywhere. There are thus only four unknowns—V, ρ, J and U.

(3·3) *Application to plane diode with uniform magnetic field parallel to plates.* Choosing the electric field parallel to the axis of x and normal to the plates, and the magnetic field along the axis of z, we note that terms in $\partial U_x/\partial y$, $\partial U_x/\partial z$, $\partial U_y/\partial y$, $\partial U_y/\partial z$ may be omitted on account of considerations of symmetry, while U_z is unaffected by either field and $\partial U_y/\partial t$ is zero in the absence of periodic electric forces in the yz plane. Thus only two of the nine* terms of $U\nabla . U$ differ from zero, and the equations required are simply

$$\frac{\partial^2 V}{\partial x^2} = -4\pi c^2 . \rho \qquad \text{......(35)},$$

$$J_x = \rho U_x - \frac{1}{4\pi c^2}\frac{\partial^2 V}{\partial x \partial t} \qquad \text{......(36)},$$

$$\frac{\partial U_x}{\partial t} + U_x \frac{\partial U_x}{\partial x} = -\frac{e}{m}\frac{\partial V}{\partial x} + \omega U_y \qquad \text{......(37)},$$

$$U_x \frac{\partial U_y}{\partial x} = -\omega U_x \qquad \text{......(38)},$$

$$\frac{\partial J_x}{\partial x} = 0 \qquad \text{......(39)}.$$

(3·4) *Derivation of primary equation.* Equation (38) gives by direct integration, after division throughout by U_x

$$U_y = -\omega x \qquad \text{......(40)}.$$

The value of U_y at $x=0$ is zero since the y component of emission velocity vanishes when averaged over all electrons. With the help of equation (40) we can eliminate U_y so that equation (37) becomes

$$\frac{\partial U_x}{\partial t} + U_x \frac{\partial U_x}{\partial x} = -\frac{e}{m}\frac{\partial V}{\partial x} - \omega^2 x \qquad \text{......(37}a\text{)}.$$

We differentiate this equation partially first with respect to t

$$\frac{\partial}{\partial t}\left(\frac{\partial U_x}{\partial t} + U_x \frac{\partial U_x}{\partial x}\right) = -\frac{e}{m}\frac{\partial^2 V}{\partial x \partial t} \qquad \text{......(41)},$$

and then with respect to x instead of t

$$\frac{\partial}{\partial x}\left(\frac{\partial U_x}{\partial t} + U_x \frac{\partial U_x}{\partial x}\right) = -\frac{e}{m}\frac{\partial^2 V}{\partial x^2} - \omega^2 \qquad \text{......(42)}.$$

Multiplying equation (42) by U_x and adding to equation (41), we obtain

$$\left(\frac{\partial}{\partial t} + U_x \frac{\partial}{\partial x}\right)\left(\frac{\partial U_x}{\partial t} + U_x \frac{\partial U_x}{\partial x}\right) = -\frac{e}{m}\left(\frac{\partial}{\partial t} + U_x \frac{\partial}{\partial x}\right)\frac{\partial V}{\partial x} - \omega^2 U_x$$

or

$$\left(\frac{\partial}{\partial t} + U_x \frac{\partial}{\partial x}\right)^2 U_x = 4\pi \frac{e}{m} c^2 J_x - \omega^2 U_x \qquad \text{......(43)}.$$

Equation (43) is the starting point for further calculations and will be referred to as

* For the x component of $U\nabla . U$ see (2.1).

the *primary equation*. It is to be noted that the primary equation contains only the unknowns U_x and J_x and we are able to impress a known value of J_x measurable outside the system. The unknowns U_x, ρ and V will all be expressible in terms of J_x once the solution for U_x is obtained. We shall thus have a relation between V and J_x, both of which are measurable* in the external circuit. The advantage of working with a known impressed current rather than a known impressed potential arises from the following consideration: when the current is known in the external circuit it is also known at all points between the plates, in view of equation (39). Every attempt to obtain a solution by working with V and J has so far proved discouraging. It is especially convenient to work with U when a magnetic field is present in view of the last term of equation (43).

§4. SOLUTION FOR THE STEADY STATE

(4.1) *Average velocity and acceleration*. We shall require equation (43) in which, for the steady state,

$$\frac{\partial}{\partial t}=0, \quad \frac{\partial}{\partial x}=\frac{d}{dx}, \quad U_x=\overline{U}, \quad J_x=\overline{J}.$$

Writing $$4\pi\frac{e}{m}c^2\overline{J}=C,$$ C

we may henceforth think of C as the *steady current*, the constant $4\pi(e/m)c^2$ being understood. Equation (43) reduces to

$$\overline{U}\frac{d}{dx}\left(\overline{U}\frac{d\overline{U}}{dx}\right)=C-\omega^2\overline{U} \qquad \text{......(44)}.$$

If T is the transit time in the steady state, we have T

$$\overline{U}\frac{d}{dx}=\frac{d}{dT}.$$

We now introduce the variable θ, which will be used throughout and is defined by θ

$$\theta=\omega T \qquad \text{......(45)}.$$

Since θ/ω represents the transit time to the plane x, we may think of θ as "x at a given ω." We may also think of θ as "ω at a given x."

Equation (44) becomes

$$\frac{d^2\overline{U}}{d\theta^2}+\overline{U}=\frac{C}{\omega^2}.$$

The solution of which is of the form

$$\overline{U}=\frac{C}{\omega^2}+A_1\cos\theta+A_2\sin\theta \qquad \text{......(46)},$$

$$\frac{d\overline{U}}{d\theta}=A_2\cos\theta-A_1\sin\theta \qquad \text{......(47)}.$$

* The question whether acceptable accuracy can be achieved with existing technique depends upon a number of considerations, such as frequency, amplitude, and ratio of alternating to direct current or voltage. In general the measurements would present difficulties.

A_1, A_2 The arbitrary constants A_1 and A_2 are completely determined by assigning values to $\bar{U}$ and $d\bar{U}/d\theta$ at the cathode, which is taken to coincide with the plane $x=0$. We only have to consider the components normal to the cathode. As the y and z components of velocity vanish at the cathode when averaged over all electrons with y and z components ranging between $-\infty$ and $+\infty$, the x component of initial velocity is also the initial velocity itself in the coordinate system now in use. This is nearly, but not quite, true of the initial acceleration as will appear immediately.

Accordingly the initial velocity is given by

$$(U_0, 0, 0) \qquad \text{......(48).}$$

The initial acceleration has the components

$$\left(\frac{e}{m}X_0, -\omega U_0, 0\right) \qquad \text{......(49),}$$

X_0 where X_0 is the electric field at the cathode. The y component of initial acceleration, which arises from the x component of velocity in view of equation (38), does not here concern us, since equations (46) and (47) refer only to the x components of

U_0 velocity and acceleration respectively. Writing $\bar{U}=U_0$, when $\theta=0$, and also

$$\frac{d\bar{U}}{d\theta}=\frac{1}{\omega}\frac{e}{m}X_0,$$

we have at the cathode

$$U_0=\frac{C}{\omega^2}+A_1 \qquad \text{......(46}a\text{),}$$

$$\frac{1}{\omega}\frac{e}{m}X_0=A_2 \qquad \text{......(47}a\text{).}$$

Then boundary conditions (48), (49) suffice to determine $\bar{U}$ everywhere in terms of C and θ. Equations (46a), (47a) determine the arbitrary constants A_1 and A_2. Accordingly equations (46) and (47) become

$$\bar{U}=\frac{C}{\omega^2}(1-\cos\theta)+U_0\cos\theta+\frac{1}{\omega}\frac{e}{m}X_0\sin\theta,$$

$$\frac{d\bar{U}}{d\theta}=\frac{C}{\omega^2}\sin\theta-U_0\sin\theta+\frac{1}{\omega}\frac{e}{m}X_0\cos\theta.$$

n_1, n_2 For convenience we define n_1 and n_2 as follows:

$$\left.\begin{aligned} n_1&=1-\frac{\omega^2 U_0}{C} \\ n_2&=\omega\left(\frac{e}{m}X_0\right)C^{-1} \end{aligned}\right\} \qquad \text{......(50).}$$

Typical values of n_1 and n_2 are 0·88 and $-0{\cdot}48$.

The quantity n_1 may be thought of as a correcting factor for finite emission velocities, and n_2 as proportional to the acceleration at the cathode. The fact that n_1 and n_2 reduce to unity and zero respectively when $\omega=0$ does not mean that the corrections disappear for zero magnetic field. The values of $\bar{U}$ and $d\bar{U}/d\theta$ in the

simplified notation are as follows:

$$\bar{U}=\frac{C}{\omega^2}(1-n_1\cos\theta+n_2\sin\theta) \qquad \ldots\ldots(51),$$

$$\frac{d\bar{U}}{d\theta}=\frac{C}{\omega^2}(n_1\sin\theta+n_2\cos\theta) \qquad \ldots\ldots(52).$$

If now $\omega\to 0$ we obtain, after writing $\theta=\omega T$,

$$\bar{U}=\frac{1}{2}CT^2+U_0+\frac{e}{m}X_0T \qquad \ldots\ldots(51a),$$

$$\frac{d\bar{U}}{dT}=CT+\frac{e}{m}X_0 \qquad \ldots\ldots(52a).$$

The quantity
$$\frac{1}{\bar{U}}\frac{d\bar{U}}{d\theta}=\frac{n_1\sin\theta+n_2\cos\theta}{1-n_1\cos\theta+n_2\sin\theta} \qquad \ldots\ldots(53)$$
will later be seen to be of fundamental importance.

The drift velocity is given by $\quad U_y=-\omega x \qquad \ldots\ldots(54).$

(4.2) *The average path of electrons.* In the steady state we may write

$$U_x=\frac{dx}{dT}, \qquad U_y=\frac{dy}{dT}.$$

By integration of equation (51) with respect to T, we obtain when $x=0$ and $T=0$:

$$x=\frac{C}{\omega^3}(\theta-n_1\sin\theta-n_2\cos\theta+n_2) \qquad \ldots\ldots(55).$$

A further integration gives, if $y=y_0$ when $\theta=0$,

$$y=y_0-\frac{C}{\omega^3}\left(\frac{\theta^2}{2}+n_1\cos\theta-n_1-n_2\sin\theta+n_2\theta\right) \qquad \ldots\ldots(56).$$

The full-line curve of figure 1 gives the shape of path for electrons emitted with negligible velocity from the cathode under conditions of space-charge limitation.

Figure 1.

The well-known cycloidal path* corresponding to zero space charge is shown dotted on the same diagram. The marked difference brought about by space charge may be understood by writing x in the form

$$x=\frac{C}{\omega^3}(\theta-n_1\sin\theta)+\frac{1}{\omega^2}\left(\frac{e}{m}X_0\right)(1-\cos\theta). \qquad \ldots\ldots(55a).$$

If we start with a very low cathode temperature, the anode current will be so small that the term in C is unimportant and may be omitted.

The maximum value of x is then seen to occur when $\theta=\pi$, the corresponding path being the dotted curve of figure 1. As the cathode temperature is raised the term in C begins to be of importance and $(e/m)\,X_0$ will become less on account of

* J. J. Thomson, *Phil. Mag.* 48, 517 (1899).

space-charge limitation. The path will then be intermediate between the dotted and full-line curves. If the cathode temperature be further raised to that required for space-charge saturation of the space, X_0 vanishes and the full-line curve obtains. Any further increase in cathode temperature makes $(e/m)\,X_0$ negative but leaves the path substantially unaltered. The full-line curve may thus be regarded as corresponding to practical conditions in which the space current, when magnetic field is absent, is less than the saturation value.

(4.3) *Critical field.* The solution for individual electrons is here of assistance as it clearly demonstrates the effect of initial velocities on the turning-point of electrons. In the steady state, equation (13) shows that the velocity of electrons with zero emission velocities is

$$\dot{x} = \frac{\ddot{x}_0}{\omega} \sin\theta + 4\pi \frac{e}{m} c^2 \frac{\bar{J}}{\omega^2} (1 - \cos\theta) \qquad \text{......(13}a\text{)},$$

in which we have written $\omega t = \theta$ (as we are entitled to do in the steady state) so as to avoid confusion which might arise from the different meanings attached to t in the respective co-ordinate systems. Since the transit time will vary slightly over individual electrons, θ/ω should be thought of as that interval of time which has elapsed since any specified set of electrons left the cathode. In § 5 at (5.1) the conception of *transit time to a given plane* is adopted as it greatly aids the understanding of the time-variable case. The given "plane" has actually a finite thickness due to initial velocity-distribution. As shown in § 4 at (4.4) the thickness of this "plane" is of the order of 10^{-3} cm. only.

If, then, $\theta = 2\pi$, we find from equation (13a) that $\dot{x} = 0$, i.e. the electrons emitted with zero velocity come to rest. If the magnetic field be of such a value that $\theta < 2\pi$ everywhere, all electrons will reach the anode, but if $\theta > 2\pi$ some or all of the electrons will be returned to the cathode. It is convenient to define that magnetic field which makes $\theta = 2\pi$ at the anode under the given conditions as *the critical field*. This definition results in an average velocity at the anode equal to the average initial velocity of electrons. It was pointed out in § 2 at (2.41) that the excursion of the electrons was affected more by the y component than by the x component of initial velocity. This conclusion was intended to apply only to the case in which space-charge is negligible. The y component of velocity will not affect the position of the anode corresponding to critical field. Thus, by integration with respect to t, equation (13a) becomes

$$x = \frac{\ddot{x}_0}{\omega^2} (1 - \cos\theta) + 4\pi \frac{e}{m} c^2 \frac{\bar{J}}{\omega^3} (\theta - \sin\theta) \qquad \text{......(57)}.$$

Since, by equation (6),

$$\ddot{x}_0 = \omega \dot{y}_0 + \frac{e}{m} X_0,$$

we see that when $\theta = 2\pi$ the value of $\dot{y}_0$ is without effect on the value of x. Writing

$$4\pi \frac{e}{m} c^2 \bar{J} = C,$$

we have for the cathode-anode separation for critical field

$$d = \frac{C}{\omega^3} 2\pi \qquad \text{......(58)},$$

or writing
$$\omega = \frac{e}{m} H_c$$ H_c

and rearranging
$$H_c = \frac{m}{e}\left(2\pi C d^{-1}\right)^{\frac{1}{3}} \qquad \text{......(58}\,a\text{)}.$$

Equation (58 a) gives the critical field in terms of the anode current and the anode-cathode separation.

By integration of equation (37 a) in § 3 with respect to x we obtain in the steady state, subject to $V=0$ when $x=0$,

$$U_x^2 = U_0^2 - 2\frac{e}{m} V - \omega^2 x^2.$$

For the anode potential corresponding to the critical condition, if

$$x=d, \quad U_x = U_0, \quad \omega = eH_c/m,$$

$$V_c = (-e/2m)\, H_c^2 d^2 \qquad \text{......(59)}.$$

Eliminating H_c between equations (58 a) and (59) we obtain the relation between space current and anode voltage corresponding to the critical condition

$$J_c = \frac{(2e/-m)^{\frac{1}{2}}\, V_c^{\frac{3}{2}}}{(2\pi C d)^2} \qquad \text{......(60)},$$

which is seen to be only $9/4\pi$ times the value without magnetic field*.

Since in the critical condition even the slowest electrons are only just brought to rest, the above reduction in anode current as compared with the conditions obtaining for zero magnetic field is not attributable to returning electrons. The reduction may be regarded as due to increased space-charge limitation. This point will be considered again in (4.5).

(4.4) *Magnetic field in excess of critical value.* The y component of velocity, or drift velocity, will still be in the same direction, but the x component of velocity will be reversed in sign for those electrons which are returning to the cathode. If a fraction κ of the outgoing electrons return to the cathode, then the space-charge κ density is increased in the ratio $(1+\kappa)$ provided we are only considering the steady state, in which electrons move back with a velocity which, at a given plane, is equal and opposite to the outgoing velocity.

Let J_c correspond to the space current flowing when the magnetic field is of J_c value just insufficient to return any electrons to the cathode. If ρ_c represent the ρ_c space-charge density under the above conditions, we have under the new conditions a nett space-charge density ρ equal to $\rho_c(1+\kappa)$. ρ

Let U_x, U_y be the velocity components under the new conditions; then for U_x, U_y outgoing electrons we have the current J_1, where

$$J_1 = \rho_c U_x.$$ J_1

For returning electrons we have the current J_2, where

$$J_2 = -\kappa \rho_c U_x.$$ J_2

* Langmuir, *Phys. Rev.* 2, 450 (1913).

J Then the resultant current J in the x direction is given by

$$J=J_1+J_2=\rho_c(1+\kappa)\,U_x=\rho U_x\frac{1-\kappa}{1+\kappa} \qquad \text{......(61).}$$

The appearance of $(1+\kappa)$ in the denominatoi signifies additional space-charge limitation of current.

Now
$$\frac{\partial^2 V}{\partial x^2}=-4\pi c^2\rho,$$

while
$$U_x\frac{\partial U_x}{\partial x}=-\frac{e}{m}\frac{\partial V}{\partial x}+\omega U_y$$

and
$$U_x\frac{\partial U_y}{\partial x}=-\omega U_x \qquad \text{......(62);}$$

after writing from equation (62), by integration,

$$U_y=-\omega x \qquad \text{......(63),}$$

we have
$$U_x\frac{\partial U_x}{\partial x}=-\frac{e}{m}\frac{\partial V}{\partial x}-\omega^2 x \qquad \text{......(64),}$$

whence
$$U_x\frac{\partial}{\partial x}\left(U_x\frac{\partial U_x}{\partial x}\right)=4\pi\frac{e}{m}c^2 J\frac{1+\kappa}{1-\kappa}-\omega^2 U_x \qquad \text{......(65),}$$

where
$$\left.\begin{aligned} J_c&=J\frac{1-\kappa}{1+\kappa}\\ C_c&=4\pi\frac{e}{m}c^2 J_c\end{aligned}\right\} \qquad \text{......(66).}$$

The solution is substantially as before, e.g.

$$U_x=\frac{C_c}{\omega^2}(1-n_1\cos\theta+n_2\sin\theta).$$

We can now determine the value of U_x for magnetic fields in excess of the critical
C_c value, since even if all the electrons are returned to the cathode, C_c is to be interpreted as the one-way current corresponding to the condition $\kappa=0$. The value of U_x will be affected by the larger value of ω, so that at any point between the plates the outgoing velocity is lessened on this account, as also is the potential. The velocity of the returning electrons is given by equation (67), but with the sign of C_c reversed.

As in (4.3) electrons emitted from the cathode with zero velocity come to rest when $\theta=2\pi$. In the case now under consideration such electrons reverse their direction when $\theta=2\pi$, and will be travelling back towards the cathode for values of θ given by

$$2\pi<\theta<4\pi.$$

The electrons of average emission velocity reverse their direction when

$$1-n_1\cos\theta+n_2\sin\theta=0 \qquad \text{......(67),}$$

which corresponds very nearly to

$$\theta=2\pi+\frac{1-n_1}{-n_2} \qquad \text{......(68).}$$

Such electrons do not arrive back at the cathode until

$$\theta = 4\pi + \frac{1 - n_1}{-n_2} \qquad \text{......(69).}$$

For a retarding field at the cathode surface n_2 will be negative. Giving n_1 and n_2 their values from equation (50), we find that

$$\frac{1 - n_1}{-n_2} = \frac{\omega U_0}{-\frac{e}{m} X_0} \qquad \text{......(70).}$$

For a magnetic field of 100 G.* and a retarding field† at the cathode of 100 V./cm.

$$\left.\begin{aligned} \omega &= 1{\cdot}77 \times 10^9 \text{ sec.}^{-1} \\ \frac{e}{m} X_0 &= -1{\cdot}77 \times 10^{17} \text{ cm./sec.}^2 \end{aligned}\right\} \qquad \text{......(71).}$$

If we take 2×10^7 cm./sec. as the mean emission velocity, equation (69) becomes

$$\theta = 2\pi + \tfrac{1}{5} \qquad \text{......(72).}$$

This corresponds to a change of some 3 per cent in θ. The maximum distance travelled in the x direction by electrons emitted with average velocity is given by $\bar{x}$, where $\bar{x}$

$$\bar{x} = \frac{C_c}{\omega^3} (\theta + n_2 - n_2 \cos\theta - n_1 \sin\theta)$$

in which $$\theta \doteq 2\pi + \frac{1 - n_1}{-n_2}.$$

Inserting the above value of θ we obtain after reduction

$$\bar{x} = \frac{2\pi C_c}{\omega^3} + \frac{U_0^2}{2(-e/m) X_0}.$$

With the values of U_0 and $(e/m) X_0$ given under equation (71), the value of the last term is only about 10^{-3} cm., which shows that all electrons turn back at substantially the same value of x given by $2\pi C_c/\omega^3$.

(4.5) *Potential-distribution.* The value of the potential $\bar{V}$ at any value of x may be obtained from the equation $\bar{V}$

$$2\left(\frac{-e}{m}\right) \bar{V} = \bar{U}^2 - \bar{U}_0^2 + \omega^2 x^2 \qquad \text{......(73).}$$

In this connection it is not convenient to express the velocity $\bar{U}$ in terms of x, $\bar{U}$
and the most satisfactory method is to express $\bar{V}$ entirely in terms of θ and obtain the $\bar{V}$, x relation by graphical means with the aid of the θ, x relation expressed by equation (55). The $\bar{V}$, θ relation is as follows:

$$\bar{V} = \frac{mC^2}{-e\omega^4} \Big[\frac{1}{2}\theta^2 - n_1 \theta \sin\theta + n_1 (1 - \cos\theta) + n_2 \{(1 - n_1)\sin\theta + (\theta + n_2)(1 - \cos\theta)\} \Big] \qquad \text{......(74),}$$

* When ω is positive, H has same sign as e/m.
† X_0 is positive for a retarding field.

while from (4.2) $$x=\frac{C}{\omega^3}(\theta+n_2-n_1\sin\theta-n_2\cos\theta) \qquad \text{......(55 }a\text{).}$$

In the special case of zero initial velocities and accelerations we obtain the following simplified relations for $\bar{V}$ and x

$$\bar{V}=\frac{mC}{-e\omega^4}\left[\frac{1}{2}\theta^2-\theta\sin\theta+1-\cos\theta\right] \qquad \text{......(74 }a\text{),}$$

$$x=\frac{C}{\omega^3}[\theta-\sin\theta] \qquad \text{......(55 }b\text{).}$$

Figure 2. Effect of magnetic field on potential-distribution (planes).

Equations (74) and (55 *a*) give the potential-distribution under any conditions of magnetic field, from zero to the critical value. The potential-distribution under the condition of critical field, from equations (74 *a*) and (55 *b*), is given by curve (1) in figure 2. For comparison with curve (1) the following curves (2) to (5) have been added to figure 2 in thin lines:

$$(2)\ V\propto x^{1\cdot5},\quad (3)\ V\propto x^{4/3},\quad (4)\ V\propto x,\quad (5)\ V\propto x^2 \qquad \text{......(75).}$$

The close correspondence of curve (1) with curve (2) indicates that the potential-distribution obeys a 1·5-power law very nearly for points not too close to the

cathode. Curve (3) corresponds to the potential-distribution for zero magnetic field, and curve (4) to that for negligible space charge, while curve (5) is purely hypothetical.

The potential-distribution in the presence of a magnetic field is found to differ so little from that for zero magnetic field that the analysis of previous workers* for the initial velocity-correction and potential-minimum may be used for magnetic fields up to 100 G. with an error of only 2 per cent in the minimum potential. In view of the above considerations a full treatment of the potential-distribution in the presence of a magnetic field is thought not to be warranted.

Following Richardson we write for the number dN_s of electrons emitted per unit area per second with forward components of velocity lying between $\dot{x}_0$ and $\dot{x}_0 + d\dot{x}_0$ N_s

$$dN_s = N_s \frac{m\dot{x}_0}{kT_c} e^{-m\dot{x}_0^2/2kT_c} \, d\dot{x}_0 \quad \text{......(76)},$$

in which the temperature T_c is provided with a suffix denoting "cathode," in order to distinguish it from T the transit time. T_c

If the current flowing to the anode is less than the saturation current (corresponding to N_s) this deficiency must be due to a retarding potential-gradient close to the surface of the cathode by which the more slowly moving electrons are forced back to the cathode. If the potential of the anode is positive and that of the cathode zero, there must be a surface of negative potential between cathode and anode at which the potential is a minimum. This conclusion also applies when a magnetic field is present: those electrons which fail to reach the potential-minimum when no magnetic field is present will still fail to do so when the magnetic field is applied.

If V_m represents the potential at the potential-minimum surface, J_s the saturation current and J the actual anode current, only those electrons which are emitted with sufficient velocity to overcome the potential-difference contribute to the anode current J. V_m, J_s J

When we take the magnetic field into account the kinetic energy of individual electrons at any point is given by

$$\tfrac{1}{2}m\dot{x}^2 = \tfrac{1}{2}m\dot{x}_0^2 - eV - \tfrac{1}{2}m\omega^2 x^2 \quad \text{......(77)}.$$

The initial energy $\frac{1}{2}m\dot{x}_0^2$ of the electrons will enable them to overcome the potential minimum if

$$\tfrac{1}{2}m\dot{x}_0^2 \geqslant eV_m + \tfrac{1}{2}m\omega^2 x_m^2,$$

in which e and V_m are both negative, so that eV_m is positive. Electrons corresponding in number to J_s are being emitted continuously from the cathode even when the current is not saturated, but a certain fraction of them are then made to return by the retarding field existing close to the cathode.

By integrating equation (76) with respect to $\dot{x}_0$ between the limits

$$(2eV_m/m + \omega^2 x_m^2)^{\frac{1}{2}}$$

* Epstein, *Ber. dtsch. phys. Ges.* **21**, 85 (1919); Fry, *Phys. Rev.* **17**, 441 (1921); Langmuir, *Phys. Rev.* **21**, 419 (1923).

and ∞ and dividing by

$$N_s\left(\frac{kT_c}{2\pi m}\right),$$

we obtain the modified Boltzmann equation

$$\frac{J}{J_s}=\epsilon^{-(eV_m+\frac{1}{2}m\omega^2x_m^2)/kT_c}=\epsilon^{-eV_m'/kT_c} \qquad \text{......(78).}$$

V_m' A little consideration shows that V_m' is in fact the new value of the minimum potential. The magnetic field affects the space charge by curving the paths of the electrons. While at points very close to the cathode the direction of motion of electrons is hardly altered, the path of electrons at points further removed from the cathode is such as to increase the space charge. This in turn depresses the potential at all points, including those very near to the cathode. The term $\frac{1}{2}m\omega^2x_m^2$ may be interpreted as electric potential energy arising from the influence of the magnetic field on the space charge.

With the help of Langmuir's analysis we then find the values of V_m and x_m corresponding to negligible magnetic field, and insert these values in equation (78) to see what effect arises from the term $\frac{1}{2}m\omega^2x_m^2$.

In illustration of the procedure let us consider an example representative of present-day practice. Instead of a bright tungsten emitter let us take a chemically coated cathode operating at 900° C., so that $T_c=1173°$ K. This is the first occasion that we have had to consider the nature of the emitting surface. The theory can be applied whatever the values of A and b, the thermionic constants of Richardson*. Provided that there is no positive emission from the cathode we are free to apply the analysis to any form of emitting surface free from irregularities.

A, b

The proportion of the total emission which can safely be taken from a chemically coated cathode varies considerably with the nature of the coating, but would not generally exceed 25 per cent for a reasonable life. In the case of a typical coating operating at 1173° K. the total emission will be around 200 mA./cm.² Let us then take J to be 50 mA./cm.² Neglecting magnetic field and following Langmuir† we find that

$$\frac{eV_m}{kT_c}=\log_\epsilon\frac{J_s}{J}=\log_\epsilon 4=1{\cdot}3863=\eta_1.$$

ξ, η; ξ_1, η_1

Langmuir's table II connecting the variables ξ, η defined by his equations (5) and (6) may then be used to determine the value ξ_1 of ξ corresponding to η_1. We thus obtain

$$-\xi_1=1{\cdot}7950.$$

From Langmuir's equations (19), (20) and (21), if we write $x_1=0$ to correspond with the cathode, we obtain

$$\left.\begin{aligned}V_m&=-0{\cdot}14022\text{ V.}\\ x_m&=1{\cdot}7753\times10^{-3}\text{ cm.}\end{aligned}\right\} \qquad \text{......(79).}$$

* *Emission of Electricity from Hot Bodies* (Longmans, Green & Co., 1921).
† *Loc. cit.*

From equation (78) we have

$$V_m' = V_m + \frac{m}{2e}\omega^2 x_m{}^2 = V_m + \frac{e}{2m} H^2 x_m{}^2 \qquad \text{......(80).}$$

Taking H as 100 G. and substituting for V_m and x_m from equation (79), we find that

$$V_m' = -0{\cdot}1402 - 0{\cdot}0027 = -0{\cdot}1429 \text{ V.},$$

thus confirming that the potential minimum is increased by the magnetic field by 2 per cent only when $H = 100$ G.

But for a magnetic field equal to the critical value we find by equation (58a) when $J = 50$ mA./cm.² and $d = 0{\cdot}5$ cm.

$$H_c = 131{\cdot}60 \text{ G.}$$

Insertion of this value of H in equation (80) shows that in the example under consideration the correction for magnetic fields up to the critical value amounts to 3·5 per cent in V_m. A reduced anode current as compared with that for zero magnetic field, as given by equation (60), must receive explanation on the basis of an increased number of electrons being unable to overcome the potential-minimum. A reduction of 28·4 per cent in J seems at first sight too great to be accounted for by a 3·5 per cent increase in $(-V_m)$, but it must be borne in mind that, since $J_s = 4J$, the corresponding change in the returning current $J_s - J$ is much less, amounting to some 14 per cent only.

(4.6) *The electric field and the initial velocity.* As compared with equation (74) for the potential, we have for the electric field the refreshingly simple relation

$$X = \frac{Cm}{\omega e}(\theta + n_2) = \frac{Cm}{\omega e}\theta + X_0 \qquad \text{......(81).}$$

Equation (81), which is equivalent to equation (20) of § 2, may be confirmed by taking $-dV/d\theta$ from equation (74) and dividing by $dx/d\theta$.

When $\theta = 2\pi$ we have for the electric field

$$X = X_0 + \frac{2\pi C}{\omega}\frac{m}{e}.$$

If $\theta = 2\pi$ at the anode we obtain with the help of equations (58) and (59), writing $X = X_c$ and eliminating C/ω,

$$X_c = X_0 - 2V_c/d \qquad \text{......(82).}$$

Equation (82) gives the electric field at the anode in the critical condition. If $X_0 \leqslant 2\dfrac{V_c}{d}$, we note that equation (82) gives a value of $(-X_c)$ 50 per cent greater than the value $\frac{4}{3}V/d$ corresponding to zero magnetic field.

Of more importance than the electric field at the anode is the electric field at the cathode. Negative at all points to the right of the potential minimum, the electric field vanishes at the potential minimum and becomes positive at the cathode. Equation (81) shows that the value θ_m of θ corresponding to the potential minimum is given by

$$\theta_m = -n_2.$$

T_m The transit time to the potential minimum is accordingly T_m, where

$$T_m = \frac{-n_2}{\omega} = \frac{-eX_0}{mC} \qquad \text{......(83).}$$

In order to determine T_m and X_0 let us write $\theta = \theta_m$, $n_2 = -\theta_m$ in equations (74) and (55 *a*). We then obtain

$$V_m = \frac{mC^2}{-e\omega^4}\left[\frac{1}{2}\theta_m{}^2 - \theta_m \sin\theta_m + n_1(1-\cos\theta_m)\right] \qquad \text{......(84),}$$

$$x_m = \frac{C}{\omega^3}(\theta_m \cos\theta_m - n_1 \sin\theta_m) \qquad \text{......(85).}$$

Eliminating n_1 between equations (84) and (85) we obtain

$$V_m = \frac{mC^2}{-e\omega^4}\left[\frac{1}{2}\theta_m{}^2 - \left(\theta_m + \frac{\omega^3 x_m}{C}\right)\left(\frac{1-\cos\theta_m}{\sin\theta_m}\right)\right] \qquad \text{......(86).}$$

When the magnetic field is negligible equations (85) and (86) reduce to

$$x_m = U_0 T_m - \tfrac{1}{3} C T_m{}^3 \qquad \text{......(85 a),}$$

$$V_m = -\tfrac{1}{2} X_0 x_m + \frac{m}{e}\frac{C^2}{24} T_m{}^4 \qquad \text{......(86 a).}$$

Now in (2.4) it was shown that for regions in which electrons are moving in two directions at once the value of C at any point corresponds to the outgoing current at that point. Now at the cathode we have a total outgoing current of 200 mA., but since electrons are returning to the cathode at all points between the cathode and the potential-minimum, the outgoing current continually decreases until at the potential-minimum itself the outgoing current becomes equal to the anode current, namely 50 mA. in the example of the previous section. The value of C corresponding to an outgoing current of 50 mA. is (if we take $\frac{C}{m} = 1{\cdot}766$ to correspond with Langmuir's figure)

$$C = 0{\cdot}9987 \times 10^{+27} \text{ cm./sec.}^3 \qquad \text{......(87).}$$

If V_m, x_m are regarded as known, there are still three unknowns in equations (85 *a*), (86 *a*), namely X_0, U_0 and T_m.

We must therefore use a third equation, e.g. (83), to eliminate X_0 from equation (86 *a*), when we shall be able to derive T_m. Having once determined T_m, we can then obtain X_0 and U_0 from equations (83) and (85 *a*) respectively.

Eliminating X_0 between equations (83) and (86 *a*)

$$2\frac{eV_m}{m} = Cx_m T_m + \frac{1}{12} C^2 T_m{}^4 \qquad \text{......(88).}$$

Inserting in the above equation the values of V_m, x_m and C given in equations (79) and (87), taking $-e/m$ equal to $1{\cdot}766 \times 10^7$, and e.m.u. per gm. multiplying V_m by 10^8 to bring it to e.m.u., we have

$$4{\cdot}9517 \times 10^{14} = 1{\cdot}7770 \times 10^{24}\, T_m + 8{\cdot}3112 \times 10^{52}\, T_m{}^4.$$

The solution of this equation may be found graphically and is

$$T_m = 2{\cdot}0154 \times 10^{-10} \text{ sec.} \qquad \text{......(89).}$$

With the above value of T_m equation (83) gives for the retarding field at the cathode surface

$$X_0 = 113{\cdot}96 \text{ V./cm.} \qquad \text{......(90)},$$

while from equation (85 *a*), which may be written

$$U_0 = \frac{x_m}{T_m} + \frac{CT_m{}^2}{3},$$

we obtain for the average emission velocity of electrons constituting the anode current

$$U_0 = 2{\cdot}2330 \times 10^7 \text{ cm./sec.} \qquad \text{......(91)}.$$

The above value of U_0 exceeds by only 0·35 per cent the value U_m defined by U_m

$$U_m = \sqrt{\left(\frac{2eV_m}{m}\right)} = 2{\cdot}2253 \times 10^7 \text{ cm./sec.} \qquad \text{......(92)},$$

so that the percentage of electrons in the anode current which have velocities greatly in excess of the minimum value required to enable them to overcome the potential-minimum must in the present case be very small. For comparison with equations (91) and (92) we have the following expression for the mean velocity of emission $\overline{U}_0$ of *all* electrons emitted from the cathode, including those which turn back to the cathode without reaching the potential-minimum

$$\overline{U}_0 = \left(\frac{\pi k T_c}{2m}\right)^{\frac{1}{2}},$$

where k is Boltzman's constant and $= 1{\cdot}372 \times 10^{-16}$ erg./deg. k

This equation gives in the present example, in which $T_c = 1173°$ K., the value

$$\overline{U}_0 = 1{\cdot}6750 \times 10^7 \text{ cm./sec.}$$

The small difference between U_0 and U_m may be associated with the exponential character of the Maxwellian law of distribution of velocities, with the fact that $\overline{U}_0$ is considerably less than U_m, and with the finite probability that electrons of velocity less than U_m will succeed in penetrating the barrier*.

The data relevant to the example, and also for the case in which $J/J_s = 0{\cdot}4$ are summarized in table 1, in which columns 3 and 11 give respectively the approximate anode voltage for negligible and critical magnetic fields. As has already been pointed out, the potential-minimum and associated data will remain sensibly unaffected when a magnetic field is applied.

Table 1. $J_s = 200$ mA./cm.² $T_c = 1173°$ K. $d = 0{\cdot}5$ cm.

1	2	3	4	5	6	7	8	9	10	11
J (…/cm.²)	$\frac{J}{J_s}$	V (V.)	$-V_m$ (V.)	$10^3 x_m$ (cm.)	$10^{10} T_m$ (sec.)	X_0 (V./cm.)	$10^{-7} U_0$ (cm./sec.)	$10^{-7} U_m$ (cm./sec.)	$10^{-7} \overline{U}_0$ (cm./sec.)	V_c (V.)
50	0·25	300	0·14022	1·7753	2·0154	113·96	2·2330	2·2253	1·6750	380
80	0·4	410	0·09266	1·1925	1·3479	121·97	1·8524	1·8095	1·6750	520

* Quantum-mechanical considerations have generally been omitted in this paper, as they are believed to have importance only very near to the cathode. The absolute value of the potential-minimum would probably be somewhat less on wave-mechanical considerations, which seem to indicate a smaller space-charge density than classical theory.

The potential-distribution near the cathode corresponding to the above data is indicated on figure 3. The values of X_0 and U_0 given in columns 7 and 8 may be regarded as typical and will be used in future examples. It is noteworthy that X_0 comes out greater when $J/J_s=0{\cdot}4$ than when $J/J_s=0{\cdot}25$. This does not prevent a greater proportion of slowly moving electrons reaching the anode, the value of the minimum potential being the deciding factor.

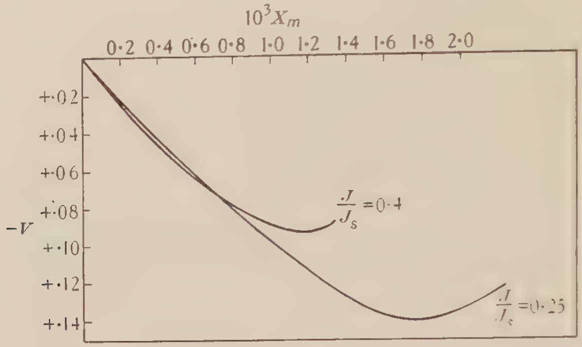


Figure 3.

(4.7) *The transit time between cathode and anode.* (4.71) *Influence of space charge.* We now come to the most important of all factors affecting the oscillations obtained in practice, whether by the magnetron method or by using a triode valve. Any error in the estimation of electron transit time might result in a completely erroneous conception of the mechanism which governs the excitation of ultra-high-frequency oscillations.

Neglecting space charge and initial velocities, Okabe* finds for the relation between transit time and critical field

$$t=\frac{\pi cm}{eH},$$

which in our present notation and units corresponds to

$$\theta_c=\pi.$$

But we have already seen (4.3) that the critical condition corresponds to

$$\theta_c=2\pi.$$

Since Okabe's result applies to both planes and cylinders in the absence of space charge effects, we cannot account for such a large difference in θ_c on the basis of electrode-geometry. Moreover, comparing Okabe's equation (9) (and putting $r'=d$ in it) with equation (59) of the present paper, we see that the relation between critical magnetic field and anode voltage is the same whether space charge is neglected or included, so that we must conclude that the difference in θ_c corresponds entirely to a difference in transit time. The 100-per-cent increase in transit time indicated by the present paper is to be regarded as a result of including the effects

* *Proc. Inst. Radio Engrs*, N.Y., **17**, 1, 652 (1929).

of space charge: it would perhaps not be altogether true to say that the increase is brought about by the space charge itself, but rather that the neglect of space charge cannot be justified except in cases where the total emission is too small to be of interest. Except for very small emission currents, equation (55) of the present paper shows that the maximum value of x is not reached by electrons until $\theta = 2\pi$. In order to check this point we may try putting θ equal to π in equation (55). We then obtain

$$x = \frac{C}{\omega^3}(\pi + 2n_2).$$

In all cases where the anode current is less than or just equal to the total emission, n_2 will be negative or zero and the value of x corresponding to $\theta = \pi$ will be less than or equal to *half* the value $2\pi C/\omega^3$ corresponding to $\theta = 2\pi$. If the cathode-temperature is reduced considerably below that corresponding to the space-charge saturation value, then $\pi < \theta_c < 2\pi$ but, as we shall see, this case is not of importance. Thus, in equation (51) for the velocity, if we write $\theta = \pi$ we find that U has the value

$$U = \frac{C}{\omega^2}(1 + n_1) = \frac{2C}{\omega^2} - U_0.$$

The condition $\theta = \pi$ could only correspond to the critical condition if the anode current were of such a value that $U = U_0$, or

$$\frac{2C}{\omega^2} = 2U_0,$$

giving

$$|J| = \frac{\omega^2 U_0}{4\pi |e/m| c^2} \times 10^4 \text{ mA./cm}^2,$$

which corresponds to 6 mA./cm². in a typical case. But we have already seen that, for $\theta = \pi$ to correspond to the critical condition, the cathode-temperature must be below the space-charge saturation value, so that we must take $J_s = J$. A total emission-density of 6 mA./cm². is too low for oscillations to occur, and the condition $\theta = \pi$ must therefore be ruled out as corresponding to a total emission-density much smaller than the generation of oscillations in practice requires. If the anode current is only slightly below the space-charge saturation value, n_2 will be slightly positive and θ_c slightly less than 2π. While oscillations in a magnetron may sometimes occur under these conditions, the degree of space-charge saturation will always be such as to make $\theta_c = 2\pi$ very nearly.

Summarizing the above, we may say that the critical condition will be given by $\theta_c = 2\pi$ in all cases where the anode current is less than the total emission, and also when the anode current is equal to the total emission and the system is in the condition of space-charge saturation. A typical transit time ($J = 50$ mA., $J_s/J = 4$) is given by putting ω equal to $2{\cdot}33 \times 10^9$, which gives

$$T = 2{\cdot}70 \times 10^{-9} \text{ sec.*}$$

* From table 1 we see that T_m, which is included in T, amounts to some 7·5 per cent of T, a somewhat large percentage having regard to the fact that $x_m/d = 0{\cdot}355$ per cent only. This state of affairs was predicted in *Phil. Mag.* 5, 641 (1928). See for example figure 2 of that paper.

(4.72) *Influence of magnetic field.* We have next to consider to what extent the transit time is increased by the magnetic field, and for this purpose we shall require to examine our knowledge of transit times in the absence of a magnetic field.

The author first gave the formula for the transit time in the plane case* and also for the cylindrical case,† neglecting initial velocities, but including the effect of space charge. His expressions for zero magnetic field and Okabe's expressions for negligible space charge give nearly equal values for T at a given anode voltage. This suggests that the modification of transit time on account of magnetic field is just about as important as the modification on account of space charge. Megaw‡ finds by graphical solution of his equation (9) that the transit time in the cylindrical case including magnetic field is increased some 36 per cent on account of space charge if R/a is of the order of 100. McPetrie§ gives a semi-graphical method of computing transit times in the absence of a magnetic field. McPetrie's study leads to the same results as Benham's formula for cylinders and provides a rapid method for direct computations of transit time. It will appear that Megaw's estimate of 36 per cent for cylinders is much below the corresponding figure for planes, i.e. 100 per cent.

If we express our transit time in terms of the anode voltage by means of equation (59), we can compare the transit times with and without magnetic field. In the
V critical condition we have, if V be the anode voltage,

$$T=\frac{2\pi}{\omega}$$

$$=\frac{2\pi d}{\sqrt{(-2Ve/m)}}.$$

If the magnetic field be removed the transit time T' corresponding to V is T' where

$$T'=\frac{3d}{\sqrt{(-2Ve/m)}}.$$

Thus $$T/T'=\tfrac{2}{3}\pi=2{\cdot}094.$$

If, however, we had neglected space charge we should have had for the transit time corresponding to critical field

$$T_1=\frac{\pi}{\omega}=\frac{\pi d}{\sqrt{(-2Ve/m)}},$$

1 in which the suffix 1 is inserted to correspond with negligible space charge. For zero magnetic field and negligible space charge the transit time is

$$T_1'=\frac{2d}{\sqrt{(-2Ve/m)}},$$

so that $$\frac{T_1}{T_1'}=\frac{\pi}{2}=1{\cdot}571.$$

* *Phil. Mag.* **5**, 653 (March 1928).

† *Phil. Mag.* **11**, 502, equation (46) (1931).

‡ *J. Instn elect. Engrs*, **72**, 326 (1933).

§ *Phil. Mag.* **16**, 284 (1933).

If we start with no space charge and no magnetic field, the effect of either, supposed to be separately applied, could be to increase the transit time in the ratio 3 to 2 approximately. Thus

$$\frac{T'}{T_1'}=1\cdot5, \qquad \frac{T_1}{T_1'}=1\cdot571.$$

The overall increase corresponding to space charge plus critical magnetic field as compared with zero space charge and zero magnetic field is

$$\frac{T}{T_1'}=\pi=3\cdot142.$$

Summarizing the information thus far available for planes and cylinders, we may construct table 2.

Table 2

	$\frac{T'}{T_1'}$	$\frac{T_1}{T_1'}$	$\frac{T}{T_1}$	$\frac{T}{T'}$	$\frac{T}{T_1'}$	Remarks
Planes	1·5	1·571	2	2·094	3·142	Dashed letters correspond to zero magnetic field, suffix 1 to zero space charge. For cylinders, R/a has been assumed to be large.
Cylinders	1·5	1·571	—	—	—	

While the transit time for cylinders including magnetic field and space charge has not yet been worked out, table 2 suggests that T/T_1 will have the value 2. Against this there is Megaw's graphically determined value $T/T_1=1\cdot36$*. Megaw's figure was obtained on the assumption of no change in the potential distribution due to the magnetic field, and would be somewhat too low on this account.

Regarding the equation $T/T_1=2$ as established on theoretical grounds for plane electrodes, we see that the period of the characteristic oscillations of a plane solid-anode magnetron must be equal to the time of a single transit between cathode and anode, and not to twice this transit time. The analysis of the present paper points to the conclusion that any electrons returning from anode to cathode contribute nothing to the negative resistance. The electrons which return to the cathode may in fact arrive with high-frequency energy sufficient to result in deleterious cathode bombardment. This would raise the cathode-temperature and emission. Megaw* has obtained this effect experimentally in the split-anode magnetron. The split-anode magnetron† is not considered in the present paper, nor has the effect of tilted magnetic field been included.

* *Loc. cit.*

† I understand from Mr Megaw that he is at present attempting an explanation of the dynatron characteristics obtained with the split-anode magnetron, taking into account the non-radial component of electric field at all points.

FOREWORD TO §§ 5 AND 6

We now come to consider in §§ 5 and 6 the ultra dynamic condition implied by the presence of rapidly alternating currents. The aim throughout these two sections is to arrive at a complete solution for the alternating-current component of electron-velocity, hereafter referred to as *alternating velocity*, in terms of the alternating current itself, and to examine the solution obtained in particular cases. The alternating current is assumed as a boundary condition; but since the value of the alternating current is specified, not only at both boundaries of the valve but also at all intermediate points, the assignment of a value to the alternating current amounts to something more than the determination of a boundary condition. As has been established in previous work*, the physical quantities potential, space-charge density and alternating velocity will all be delayed in phase with respect to the alternating current, which includes both electron and displacement current. It is, of course, possible to work out the electron current separately, and when this is done the potential is seen to lead the electron current in phase. From the practical point of view no means of distinguishing between electron and displacement current is available, since these combine at the boundaries of the valve. To be able to express the alternating velocity and potential in terms of the total alternating current is thus of paramount importance from the practical viewpoint, and in order to see whether work is done on or by the electrons all that is necessary is to multiply together the in-phase components of current and potential, and if the sign of the coefficient of $\sin^2 pt$ thus obtained is negative, power will be expended on the electrons by the alternating forces, and oscillations may result. Since the in-phase component of current is assumed positive, it is not even necessary to take the product of the in-phase components, but merely to examine the sign of the coefficient of $-\sin pt$ in the potential to see if this is ever positive. Since the quadrature component of current is zero by definition, the quadrature component of potential contributes no power either positive or negative, when averaged over a complete cycle.

The mathematics of §§ 5 and 6 require a knowledge of partial differential equations and of circular functions, but nothing else of an advanced nature. A word of warning may not be out of place to those attempting the details of the calculations, some of which are the result of a considerable amount of work. (1) No shortening of the work is possible by recourse to complex notation, and the chance of a mistake is thereby considerably increased. (2) The process of solution is rendered more difficult if the magnetic field is taken as zero. The result for zero magnetic field may always be obtained by letting ω tend to zero in the final solution. (3) In the case $U_0 = 0$, $X_0 = 0$ alternative methods of arriving at the general solution may suggest themselves. The method given in the text for the solution of equation (98) has the advantage of being strictly orthodox.

* Megaw, *loc. cit.*

§ 5. GENERAL SOLUTION

(5.1) *Reduction of primary equation to linear form.* No method has yet been found for solving directly for U_x, equation (43), except (4.1) in the steady state. It is, of course, possible to find variables τ, t_0 such that

$$\left.\frac{\partial}{\partial t}\right|_x + U_x \left.\frac{\partial}{\partial x}\right|_t = \left.\frac{\partial}{\partial \tau}\right|_{t_0}.$$

But in this case τ is the instantaneous transit time and the solution obtained will be the same as that for individual electrons, in which the arguments of the circular functions all have a fluctuating component at a given value of x. Now we are particularly anxious to avoid this feature, and we therefore seek to express equation (43) in terms of variables of which one shall be independent of time at a given value of x. Now equation (43) as it stands does satisfy this condition, since one of the independent variables is x itself. That we seek a change of variables from x and t is due entirely to the fact that the operator $(U_x\, \partial/\partial x)$ does not apply when t is constant, since U_x is itself a function of t; alternatively if we regard $\left.\frac{\partial}{\partial x}\right|_t$ as the operator, we are faced with a partial differential equation of the third degree in U_x. τ

The only method available for achieving the desired results is that of harmonic subdivision already described in (2.1) and (2.3). We shall then merely have to find variables T, t_0 such that

$$\left.\frac{\partial}{\partial t}\right|_x + \bar{U} \left.\frac{\partial}{\partial x}\right|_t = \left.\frac{\partial}{\partial T}\right|_{t_0} \qquad \text{......(93)},$$

a relatively simple matter, since $\bar{U}$ is independent of t and T will therefore be the steady state value of the transit time to the plane x. We can in fact replace $\bar{U}\,\partial/\partial x$ by $\partial/\partial T$ so that equation (93) becomes

$$\left.\frac{\partial}{\partial t}\right|_T + \left.\frac{\partial}{\partial T}\right|_t = \left.\frac{\partial}{\partial T}\right|_{t_0} \qquad \text{......(94)}.$$

The following analysis suffices to show that t_0 represents the instant at which electrons leave the cathode. Let f be any function of t, T and let $\bar{f}$ be the same function when expressed in terms of t_0, T'. $f, \bar{f}$

Then
$$\frac{\partial f}{\partial t}\delta t + \frac{\partial f}{\partial T}\delta T = \frac{\partial \bar{f}}{\partial t_0}\delta t_0 + \frac{\partial \bar{f}}{\partial T'}\delta T'.$$

Writing
$$\delta T = \frac{\partial t}{\partial t_0}\delta t_0 + \frac{\partial t}{\partial T'}\delta T',$$

we find that
$$\frac{\partial f}{\partial t}\left(\frac{\partial t}{\partial t_0}\delta t_0 + \frac{\partial t}{\partial T'}\delta T'\right) + \frac{\partial f}{\partial T}\delta T = \frac{\partial \bar{f}}{\partial t_0}\delta t_0 + \frac{\partial \bar{f}}{\partial T'}\delta T'.$$

Since $\delta T' = \delta T$ we have, comparing terms in δT,

$$\frac{\partial f}{\partial t}\frac{\partial t}{\partial T'} + \frac{\partial f}{\partial T} = \frac{\partial \bar{f}}{\partial T'}.$$

The function f is seen to satisfy equation (94) provided

$$\frac{\partial t}{\partial T'}=1,$$

which gives on integrating partially with respect to T, t_0 being constant

$$t=T'+[t_0],$$

or writing $T'=T$ we obtain $$t_0=t-T.$$

Since $T=0$ at the cathode, t_0 represents the instant at which the electrons left the cathode. Thus one of our independent variables is now the same as in (2.4), namely t_0. The other independent variable T differs from the $(t-t_0)$ of (2.4) in that the latter represents instantaneous transit time: at a given value of x the $(t-t_0)$ of (2.4) contains a fluctuating component, whereas T is constant. We may thus think of T as x itself, and no confusion will arise if we refer to the plane x as *the plane* T, bearing in mind the slight spread in T_1, see (4.3).

We shall require to replace U_x by $\bar{U}+u+\ldots$ in equation (43), which gives

$$\left[\frac{\partial}{\partial t}+(\bar{U}+u+\ldots)\right]^2(\bar{U}+u+\ldots)=4\pi\frac{e}{m}c^2J-\omega^2(\bar{U}+u+\ldots).$$

Applying equations (93) and (94) we obtain

$$\left[\left(1+\frac{u}{\bar{U}}\right)\frac{\partial}{\partial T}\bigg|_{t_0'}-\frac{u}{\bar{U}}\frac{\partial}{\partial t}\bigg|_{T}\right]^2(\bar{U}+u)=4\pi\frac{e}{m}c^2J-\omega^2(\bar{U}+u)\quad\ldots\ldots(95).$$

ψ_1 Let us extract all terms containing u once only. Let the sum of these be ψ_1. Then, since $\partial\bar{U}/\partial t=0$, we have

$$\psi_1=\frac{\partial^2 u}{\partial T^2}\bigg|_{t_0'}+\left(\frac{1}{\bar{U}}\frac{d\bar{U}}{dT}\right)\frac{\partial u}{\partial T}\bigg|_{t_0'}+\left\{\frac{2}{\bar{U}}\frac{d^2\bar{U}}{dT^2}-\frac{1}{\bar{U}^2}\left(\frac{d\bar{U}}{dT}\right)^2+\omega^2\right\}u\quad\ldots\ldots(96).$$

ψ_2 We now assume a known value for J of the type $J=\bar{J}+\sum_1^r j_n\sin^n pt$. If ψ_2 denote

ψ_3 the sum of all terms containing u twice and over ψ_3 the sum of all terms containing u three times over, equation (93) may be resolved into discrete equations of zero, first, second, third, ... order respectively

$$\frac{d^2\bar{U}}{dT^2}+\omega^2\bar{U}=4\pi\frac{e}{m}c^2\bar{J}\quad\ldots\ldots(44a),$$

$$\psi_1=4\pi\frac{e}{m}c^2 j_1\sin p\,(t_0'+T)\quad\ldots\ldots(97),$$

$$\psi_1'+\psi_2=4\pi\frac{e}{m}c^2 j_2\sin^2 p\,(t_0'+T),$$

$$\psi_1''+\psi_{12}'+\psi_3=4\pi\frac{e}{m}c^2 j_3\sin^3 p\,(t_0'+T),$$

$$\left[\begin{array}{l}\psi_1'''+\psi_{12}''+\psi_{22}'+0=4\pi\dfrac{e}{m}c^2 j_4\sin^4 pt\\ \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots\end{array}\right],$$

$, u', u''$ ψ_1', ψ_1'' correspond to the second and third-order velocities u', u'' which, for the

sake of clearness, were omitted from equation (95). ψ_n' represents the sum of third-order terms containing u once and u' once. There is no difficulty in obtaining ψ_2, ψ_2' and ψ_3 but, as was pointed out in (2.3), we shall only require the fundamental. ψ_n'

Writing $\omega T=\theta$ and giving $\bar{U}$ its value from equation (51), we obtain from equations (96) and (97) the equation

$$\frac{\partial^2 u}{\partial\theta^2}+\sigma_1\frac{\partial u}{\partial\theta}+\frac{1-n_1^2-n_2^2}{(1-n_1\cos\theta+n_2\sin\theta)^2}u=\frac{\beta}{\omega^2}\sin\left(pt_0+\frac{\theta}{\omega}\right) \qquad \ldots\ldots(98),$$

in which

$$\left.\begin{aligned}&\frac{\partial}{\partial\theta}\equiv\left.\frac{\partial}{\partial\theta}\right|_{t_0}\\ &\sigma_1\equiv\frac{1}{\bar{U}}\frac{d\bar{U}}{d\theta}=\frac{n_1\sin\theta+n_2\cos\theta}{1-n_1\cos\theta+n_2\sin\theta}\\ &\beta\equiv4\pi\frac{e}{m}c^2j_1\end{aligned}\right\} \qquad \ldots\ldots(99).$$

σ_1 β

(5.2) *Derivation of general solution.* In attempting to solve equation (98) we are assisted by our knowledge of simpler cases, which prompts us to notice that σ_1 is itself a solution to the equation when $\beta=0$. When any one solution has been obtained for the equation for $\beta=0$, it is a simple matter to obtain the general solution for $\beta=0$. This is found to be

$$\left.\begin{aligned}&A\sigma_2+B\sigma_1\\ \text{where}\qquad &\sigma_2\equiv\frac{n_1^2+n_2^2-n_1\cos\theta+n_2\sin\theta}{1-n_1\cos\theta+n_2\sin\theta}\end{aligned}\right\} \qquad \ldots\ldots(100).$$

σ_2

The general solution in case $\beta\neq 0$ then proceeds as follows. Let us change the variable to η where

$$u=\sigma_1\eta \qquad \ldots\ldots(101).$$

Equation (98) becomes, after being divided through by σ_1,

$$\frac{\partial^2\eta}{\partial\theta^2}+\left(\sigma_1+\frac{2}{\sigma_1}\frac{d\sigma_1}{d\theta}\right)\frac{\partial\eta}{\partial\theta}=\frac{\lambda}{\sigma_1}\sin(pt_0+r\theta) \qquad \ldots\ldots(102),$$

where

$$r\equiv\frac{p}{\omega},\quad \lambda\equiv\frac{\beta}{\omega^2} \qquad \ldots\ldots(103).$$

r, λ

Since equation (102) contains no term in η, the integrating factor is

$$e^{\int\left(\sigma_1+\frac{2}{\sigma_1}\frac{d\sigma_1}{d\theta}\right)d\theta}\ \text{which}=\sigma_1^2\,e^{\int\sigma_1 d\theta}=\sigma_1^2\,(1-n_1\cos\theta+n_2\sin\theta).$$

Solving for $\partial\eta/\partial\theta$ we obtain, writing $A\,(n_1^2+n_2^2)$ for the constant of integration, A

$$\frac{\partial\eta}{\partial\theta}=\frac{1-n_1\cos\theta+n_2\sin\theta}{(n_1\sin\theta+n_2\cos\theta)^2}\left[A\,(n_1^2+n_2^2)+\lambda\int(n_1\sin\theta+n_2\cos\theta)\sin(pt_0+r\theta)\,d\theta\right] \qquad \ldots\ldots(104).$$

The integral appearing in equation (104) presents no difficulty, but the next stage will involve rather unusual integrals. These have, however, given little trouble and can all be evaluated in circular functions. For example

$$\int\frac{1-n_1\cos\theta+n_2\sin\theta}{(n_1\sin\theta+n_2\cos\theta)^2}\,d\theta=\frac{N^2-n_1\cos\theta+n_2\sin\theta}{N^2\,(n_1\sin\theta+n_2\cos\theta)}.$$

σ_3 Writing σ_3 for the quantity $\dfrac{N^2-n_1\cos\theta+n_2\sin\theta}{n_1\sin\theta+n_2\cos\theta}$, we also require

$$\frac{N^2}{r-1}\int\frac{1-n_1\cos\theta+n_2\sin\theta}{(n_1\sin\theta+n_2\cos\theta)^2}\sin\left[pt_0+(r-1)\theta\right]d\theta$$
$$=\frac{\sigma_3\sin(pt-\theta)}{r-1}-\int\sigma_3\cos\left[pt_0+(r-1)\theta\right]d\theta$$

and

$$\int\sigma_3(n_1\sin\theta+n_2\cos\theta)\sin(pt_0+r\theta)\,d\theta=\int(N^2-n_1\cos\theta+n_2\sin\theta)\sin(pt_0+r\theta)\,d\theta,$$

N where $$N^2\equiv n_1{}^2+n_2{}^2,\qquad \sigma_3=\frac{\sigma_2}{\sigma_1},\qquad t=t_0+\frac{r\theta}{p}\qquad\text{......(105).}$$

With the help of the above integrals we arrive at the equation

$$\eta=A\sigma_3+B+\frac{\lambda}{r}\cos pt+\frac{\lambda}{2N^2}\left[(n_1\sigma_3-n_2)\left\{\frac{\sin(pt-\theta)}{r-1}-\frac{\sin(pt+\theta)}{r+1}\right\}\right.$$
$$\left.-(n_2\sigma_3+n_1)\left\{\frac{\cos(pt-\theta)}{r-1}+\frac{\cos(pt+\theta)}{r+1}\right\}\right],$$

which simplifies to $$\eta=A\sigma_3+B-\frac{\lambda}{r^2-1}\left[\frac{\sin pt}{\sigma_1}+\frac{\cos pt}{r}\right]\qquad\text{......(106).}$$

To obtain u we multiply by σ_1, equation (101). Giving λ and r their values from equation (103), we obtain the following general solution for u:

$$u=A\sigma_2+B\sigma_1-\frac{\beta}{p^2-\omega^2}\left[\sin pt+\frac{\omega}{p}\sigma_1\cos pt\right]\qquad\text{......(107),}$$

in which A and B are arbitrary functions of t_0 and σ_1, σ_2 are given by equations (99) and (100) respectively.

(5.3) *Notes on the general solution.* It is immaterial whether we regard u as a function of θ and t_0 or of θ and t or of all three variables θ, t_0 and t. We shall in any case retain θ, inasmuch as σ_1 and σ_2 are expressed in terms of θ. Comparison with equation (14) reveals that the factor $(p^2-\omega^2)$ is common to both solutions. In other respects there is little obvious resemblance between equations (14) and (107). The solutions are nevertheless equivalent, the large apparent difference arising from the different meaning attached to t in the two coordinate systems, as has been pointed out in (5.1).

We note that the form of equation (107) is similar to that obtained in the very much simpler case of zero initial velocities, accelerations, and magnetic field*. To obtain this case all we need to do is to write

$$\left.\begin{aligned}&\sigma_2=1,\qquad \sigma_1=\frac{\sin\theta}{1-\cos\theta},\\ &\underset{\theta\to 0}{\mathrm{Lt}}\left(\frac{\omega}{p}\sigma_1\right)=\frac{2\omega}{p\theta}=\frac{2}{pT}=\frac{2}{\xi}\end{aligned}\right\}\qquad\text{......(108).}$$

so that

* *Phil. Mag.* 5 (1928), equation (19), p. 649.

§ 6. BOUNDARY CONDITIONS

(6.1) *General.* It is usual at this stage to declare that the mathematics of the problem is at an end, for the determination of arbitrary constants and of arbitrary functions is governed by purely physical considerations. The determination of A and B in equation (107) is, however, not devoid of mathematical interest. Inspection of equation (107) reveals the existence of singularities when $p = \pm\omega$ and also when $p = 0$. It will be seen that these singularities can be removed by choosing the arbitrary functions in such a way that infinities do not occur in u either when $p = \pm\omega$ or when $p = 0$.

Any suggestion that an infinity might occur when $p = \omega$ is refuted by working out the solution with p equal to ω at the outset, when it is found that no sign of resonance occurs. Experiment, moreover, confirms this point. In general nothing unusual occurs experimentally either when $p = \omega$ or when $p = 0$, unless θ has certain values. We therefore seek to remove the singularities for all values of θ, including those which are favourable to electronic oscillations. For even in this case the alternating component of velocity cannot be infinite though, as we shall see, it may assume large values.

The number of physical conditions we shall need will not exceed the number of arbitrary constants, provided such conditions are truly independent. Now it is an experimental fact that reversal of the current in the field coil of the electromagnet is without effect on the behaviour of a magnetron, apart from experimental errors such as slight lack of alignment, and we may thus introduce as a condition governing A and B the assertion that the solution for u remains unchanged when ω is changed to $-\omega$. We cannot then, however, regard the condition $p = \pm\omega$ as providing two independent physical conditions. The condition $p = 0$ is an independent physical condition.

We have thus far three physical conditions wherewith to determine A and B. But the most general form of A and B compatible with the problem involves four arbitrary constants:

$$\left.\begin{aligned} A &= A_1 \sin pt_0 + A_2 \cos pt_0 \\ B &= B_1 \sin pt_0 + B_2 \cos pt_0 \end{aligned}\right\} \qquad \text{......(109).}$$

We shall accordingly require a fourth condition in order to determine A and B completely, unless it should so happen that A_1, A_2, B_1, B_2 are in some way related, so that they cannot all be regarded as arbitrary. While it will eventually appear that simple relations exist between the A's and the B's, there is not sufficient evidence at the present stage to assume that fewer than four arbitrary constants exist. We therefore have to seek a further physical condition which must also be independent of the three already obtained. Now while the solution must certainly remain finite when $\omega = 0$, this condition is already satisfied in equation (107). Suppose, however, it should happen that in the satisfaction of other conditions terms were introduced with singularities at $\omega = 0$; we should be justified in adjusting such terms in accordance with the condition $\omega = 0$. It will appear that this state of affairs does arise, and it does seem probable that the condition $\omega = 0$ is of paramount importance.

Now the acceleration of electrons must never be infinite, but in regard to the avoidance of infinities, care must be exercised in ensuring that any conditions imposed shall be truly independent. The question we have now to consider is whether the condition that the velocity shall never be infinite automatically excludes the possibility of an infinite acceleration. This brings us back to the old enigma of the irresistible force acting on the immovable body. Now whatever happens in this case, the irresistible force could be achieved by causing a finite momentum to suffer a change in an infinitesimal time. If, moreover, the change in momentum be brought about by a change of density rather than by a change of velocity, the velocity must remain finite, since by hypothesis it was finite before the change took place.

More than sufficient has been said to establish the fact that an exceedingly high value of acceleration could occur with an exceedingly low value of velocity, and we therefore see that by applying our conditions to the electron-acceleration the number of available physical conditions is doubled.

When the electrons leave a hot cathode they do so with a distribution of velocities, but whatever the velocity of an electron at the instant of emergence the alternating component of velocity must then be zero. Now the strange thing about this condition is the fact that it has been found of no assistance whatever in the determination of the arbitrary functions in the general case now under consideration. All attempts to arrive at the final solution on the basis that u must vanish at the hot cathode proved unavailing; yet, when the final solution was obtained on the basis of conditions connected with the avoidance of infinities in u, the condition $u=0$ at the cathode was found on inspection to be satisfied. The explanation of this apparent anomaly is as follows. The condition $u=0$ at the cathode is actually inherent in the analysis owing to the non-linearity of the primary equation, together with the introduction at an early stage of the assumed boundary values of d.-c. velocity and acceleration and of the total current. Llewellyn in a recent paper* has proposed a general form of boundary condition for u in the simple case where emission velocities are assumed to be zero. One object of Llewellyn's extra condition imposed on u was to enable the analysis to be extended to cover the case of virtual cathodes, where there is strong experimental evidence that the alternating velocity and acceleration must be very different from zero. It appears, however, from the present paper that the value of u is completely determined at a virtual cathode, no less than at a real one, without the imposition of any conditions on u, apart from those connected with the avoidance of singularities. Llewellyn's paper should be consulted for the opposite point of view, which also appears at first sight to be the natural and correct one to adopt.

Summarizing the above discussion we may say that any boundary conditions imposed on u, other than the following, are irrelevant: (i) when $p=\pm\omega$ the solution remains finite at all values of T; (ii) when ω is changed to $-\omega$ the solution remains unchanged; (iii) when $p=0$ the solution remains finite at all values of T; (iv) when $\omega=0$ the solution remains finite at all values of T.

* *Proc. Inst. Radio Engrs*, N.Y., **21**, 1532 (1933).

We also have corresponding conditions imposed on the acceleration, but it appears that A and B are completely determined by conditions (i) to (iv) above.

(6.2) *Determination of arbitrary functions.* Inspection of the general solution, e.g.

$$u = A\sigma_2 + B\sigma_1 - \frac{\beta}{p^2-\omega^2}\sin pt - \frac{\beta}{p^2-\omega^2}\frac{\omega}{p}\sigma_1 \cos pt \quad \text{......(107a)},$$

reveals that the last term has singularities when $p=0$ and also when $p=\omega$. It is therefore desirable to make use of the identity

$$\frac{\omega}{p} = \frac{p}{\omega} - \frac{p^2-\omega^2}{\omega p} \quad \text{......(110)},$$

so that we obtain

$$u = A\sigma_2 + B\sigma_1 - \frac{\beta}{p^2-\omega^2}\sin pt - \frac{\beta}{p^2-\omega^2}\frac{p}{\omega}\sigma_1 \cos pt + \frac{\beta\sigma_1}{\omega p}\cos pt \quad \text{......(107b)}.$$

We have now separated out the singularity occurring when $p=0$, so that the one term which stands in need of condition (iii) does not now require condition (i). It is true that in effecting such a separation we have produced two terms with singularities when $\omega=0$, whereas none existed before. We know, however, that these two terms balance one another, so that all we have to remember is that any values assigned to A and B must not lead to a singularity in $A\sigma_2 + B\sigma_1$ when $\omega=0$.

Condition (iii): Concerning ourselves with the removal of the singularity which occurs in the last term when $p=0$, we find that the necessary and sufficient condition that u shall remain finite when $p=0$ is that the arbitrary function B shall contribute

$$-\frac{\beta}{\omega p}\cos pt_0{}^* \quad \text{......(111)}.$$

Condition (i): Concerning ourselves next with the removal of the singularities in the two terms of equation (107b) having $p^2-\omega^2$ in the denominator, we have to give σ_1 and σ_2 their values from equations (99) and (100) and select the coefficient of $\sin pt_0$ and also the coefficient of $\cos pt_0$. Since for u to remain finite when $p=\pm\omega$, $u\,(p^2-\omega^2)$ must vanish, the above-mentioned coefficients are respectively equated to zero for the case $p=\pm\omega$.

Since factors of the type $(p/\pm\omega)^q$ may be introduced with impunity without affecting condition (i) *per se*, we have to decide the value of q with the assistance of conditions (ii), (iii) and (iv). We thus find that A and B contribute as follows:

$$\left.\begin{aligned} A\text{'s contribution} &= \frac{\beta}{p^2-\omega^2}\frac{1}{N^2}\left(-n_1 \sin pt_0 + n_2\frac{p}{\omega}\cos pt_0\right) \\ B\text{'s contribution} &= \frac{\beta}{p^2-\omega^2}\frac{1}{N^2}\left(n_2 \sin pt_0 + n_1\frac{p}{\omega}\cos pt_0\right) \end{aligned}\right\} \quad \text{......(112)}.$$

The quantities σ_1 and n_2 change sign when ω changes sign, while σ_2 and n_1 remain unchanged. Let us now confirm that condition (ii) is satisfied, i.e. that the solution remains unaffected when ω is changed to $-\omega$.

* It must be noted that whereas T is always a finite interval, t_0 and t may approach infinity since the choice of time-origin is arbitrary. In taking limits when $p\to 0$ we must therefore always leave $\sin pt$ and $\cos pt$ intact.

We have, if equations (111) and (112) represent the entire contributions of A and B:

$$u=\frac{\beta}{p^2-\omega^2}\frac{1}{N^2}\left(-n_1\sin pt_0+n_2\frac{p}{\omega}\cos pt_0\right)\sigma_2$$
$$+\frac{\beta}{p^2-\omega^2}\frac{1}{N^2}\left(n_2\sin pt_0+n_1\frac{p}{\omega}\cos pt_0\right)\sigma_1$$
$$-\frac{\beta}{p^2-\omega^2}\left(\sin pt+\frac{p}{\omega}\sigma_1\cos pt\right)+\frac{\beta}{\omega p}\sigma_1(\cos pt-\cos pt_0)\quad\ldots\ldots(113).$$

Changing ω to $-\omega$, n_2 to $-n_2$ and σ_1 to $-\sigma_1$, we obtain

$$u=\frac{\beta}{p^2-\omega^2}\frac{1}{N^2}\left[\left(-n_1\sin pt_0-n_2\frac{p}{-\omega}\cos pt_0\right)\sigma_2-\left(n_2\sin pt_0+n_1\frac{p}{\omega}\cos pt_0\right)(-\sigma_1)\right]$$
$$-\frac{\beta}{p^2-\omega^2}\left(\sin pt+\frac{p}{-\omega}(-\sigma_1)\cos pt\right)+\frac{\beta(-\sigma_1)}{-\omega p}(\cos pt-\cos pt_0),$$

showing that u remains unaltered.

We now confirm that condition (iv) is satisfied in equation (113). There are four terms in equation (112) with singularities when $\omega=0$.

We have

$$\operatorname*{Lt}_{\omega\to 0}(n_2)=0,$$
$$\operatorname*{Lt}_{\omega\to 0}(n_1)=1=\operatorname*{Lt}_{\omega\to 0}(N^2),$$

whence
$$\left.\begin{aligned}&\operatorname*{Lt}_{\omega\to 0}\left[\frac{\beta\sigma_1}{p^2-\omega^2}\frac{n_1}{N^2}\frac{p}{\omega}\cos pt_0-\frac{\beta}{\omega p}\sigma_1\cos pt_0\right]=0\\ &\operatorname*{Lt}_{\omega\to 0}\left[\frac{-\beta\sigma_1}{p^2-\omega^2}\frac{p}{\omega}\cos pt+\frac{\beta}{\omega p}\sigma_1\cos pt\right]=0\end{aligned}\right\}\quad\ldots\ldots(114).$$
and

Thus condition (iv) is satisfied. But since in arriving at equations (111) and (112) all four conditions have been used, A and B must be completely determined, and equation (113) must be the final solution for u. In view of equation (110), equation (113) may now be written

$$u=\frac{\beta}{p^2-\omega^2}\left[-\sin pt-\frac{n_1\sigma_2-n_2\sigma_1}{N^2}\sin pt_0+\frac{\omega}{p}\sigma_1(\cos pt_0-\cos pt)+\frac{p}{\omega}\frac{n_2\sigma_2}{N^2}\cos pt_0\right.$$
$$\left.-\frac{(N^2-n_1)}{N^2}\frac{p}{\omega}\sigma_1\cos pt_0\right]\quad\ldots\ldots(113a).$$

§7. DISCUSSION OF FINAL SOLUTION

(7.1) *General.* For purposes of comparison with simpler cases it is convenient to separate out terms correcting for initial velocities and accelerations. We then obtain

$$u=\frac{\beta}{p^2-\omega^2}\left[-\sin pt-\sigma_2\sin pt_0+\frac{\omega}{p}\sigma_1(\cos pt_0-\cos pt)\right]$$
$$+\frac{\beta}{p^2-\omega^2}\left\{\frac{(N^2-n_1)\sigma_2+n_2\sigma_1}{N^2}\sin pt_0-\frac{p}{\omega}\frac{(N^2-n_1)\sigma_1-n_2\sigma_2}{N^2}\cos pt_0\right\}\quad\ldots\ldots(113b).$$

In the event of initial velocities and accelerations being zero the terms between curly brackets reduce to zero, while the terms between square brackets remain. Except for very small values of θ we may write

$$\sigma_1 \doteqdot \frac{\sin\theta}{1-\cos\theta}; \quad \sigma_2 \doteqdot 1.$$

We thus obtain for zero initial velocities and accelerations

$$u=\frac{\beta}{p^2-\omega^2}\left[-\sin pt-\sin pt_0+\frac{\omega}{p}\frac{\sin\theta}{1-\cos\theta}(\cos pt_0-\cos pt)\right] \quad \text{......(113}c\text{)},$$

which is seen to correspond with equation (20) of my 1928 paper, but with $\frac{2}{\xi}$ replaced by $\frac{\omega}{p}\frac{\sin\theta}{1-\cos\theta}$. Actually we shall see immediately that $\sigma_2<1$ at points very near the cathode, so that for this reason alone equation (113*c*) cannot be applied near the cathode.

As a searching test on the uniqueness of equation (113*b*) let us seek the value of u at the cathode. We have to let T tend to zero, so that

$$\underset{T\to 0}{\mathrm{Lt}}\,(\sigma_2)=\frac{N^2-n_1}{1-n_1}=-n_1+\frac{n_2{}^2}{1-n_1} \quad \text{......(115)},$$

$$\underset{T\to 0}{\mathrm{Lt}}\,(\sigma_1)=\frac{n_2}{1-n_1} \quad \text{......(116)}.$$

When t_0 is equated to t equation (113*b*) becomes

$$\begin{aligned} u=&\frac{\beta}{p^2-\omega^2}\left[-\sin pt-\frac{N^2-n_1}{1-n_1}\sin pt\right]\\ &+\frac{\beta}{p^2-\omega^2}\left\{\frac{(N^2-n_1)^2+n_2{}^2}{N^2(1-n_1)}\sin pt-\frac{p}{\omega}\frac{(N^2-n_1)\,n_2-(N^2-n_1)\,n_2}{N^2(1-n_1)}\cos pt\right\}\\ =&\frac{\beta}{p^2-\omega^2}\left[\frac{-N^2+N^2n_1-N^4+n_1N^2+N^4-2N^2n_1+n_1{}^2+n_2{}^2}{N^2(1-n_1)}\right]\sin pt+[0]\cos pt\\ =&\,0. \end{aligned}$$

We thus see that the inclusion of initial velocities and accelerations provides all the boundary conditions required at the cathode: the value of u, the alternating velocity, reduces to zero at the cathode without the imposition of any such boundary condition on u. All attempts to arrive at equation (113) with $u=0$ as one of the four boundary conditions met with no success.

From consideration of equation (115) it is seen that the value of σ_2 at the cathode is approximately -1. It is exactly -1 at the potential-minimum. This is an interesting state of affairs which corresponds to a change of sign in σ_2, where

$$\sigma_2=\frac{N^2-n_1\cos\theta+n_2\sin\theta}{1-n_1\cos\theta+n_2\sin\theta}$$

at some point which must lie close to the cathode. Under these circumstances, giving n_1, n_2 and N^2 their values from equations (50) and (105) and writing $\theta=\omega T$,

we obtain when ωT is small

$$\sigma_2 = \frac{\left(\frac{eX_0}{mC}\right)^2 - \frac{U_0}{C} + \frac{eX_0}{mC}T + \frac{T^2}{2}}{\frac{U_0}{C} + \frac{eX_0}{mC}T + \frac{T^2}{2}}.$$

T_m Writing T_m for the time taken by electrons to travel from cathode to potential-minimum, we have in accordance with equation (83):

$$-\frac{eX_0}{mC} = T_m,$$

so that

$$\sigma_2 = \frac{T_m^2 - U_0/C - T_m T + \frac{1}{2}T^2}{U_0/C - T_m T + T^2/2} \qquad \text{......(117)},$$

which has its minimum value (-1) when $T = T_m$, i.e. at the potential-minimum. If, however, $\sigma_2 = 0$ we obtain

$$T = T_m + \sqrt{\left(\frac{2U_0}{C} - T_m^2\right)},$$

so that it is not at the potential-minimum that σ_2 changes sign but at some point further removed from the cathode, probably nearly at $x = 2x_m$. If the potential-minimum coincides with the cathode, then the approximate solution equation (113) applies at the cathode with $-\sin pt_0$ replaced by $+\sin pt_0$, thus removing any doubt that u may not vanish at the cathode according to equation (113*c*). Actually it may be regarded as somewhat remarkable that equation (113*c*) as it stands does vanish at the cathode, for if we place $t = t_0$ straight away we obtain

$$u_0 = -\frac{\beta}{p^2 - \omega^2} 2 \sin pt_0 \qquad \text{......(118)}.$$

On taking limits we find, however, a contribution from the terms in $\cos pt$, $\cos pt_0$ which exactly balance equation (118) if σ_1 is taken as $2/\theta$ at the cathode. This, in turn, is now seen to be impermissible since, writing σ_1 in full, we have when ω is small

$$(\omega\sigma_1) = \frac{\frac{e}{m}X_0 + CT}{U_0 + \frac{e}{m}X_0 T + \frac{1}{2}CT^2} \qquad \text{......(119)}.$$

Only if we write $X_0 = 0 = U_0$ *before* equating T to zero can we obtain $\sigma_1 = 2/\omega T$, and clearly the true value of σ_1 at the cathode is

$$(\sigma_1)_0 = \frac{eX_0}{m\omega U_0}.$$

For the case of zero magnetic field, for which σ_1 becomes infinite, the interesting uantities in the solution for u are σ_2 and $\underset{\omega \to 0}{\mathrm{Lt}}(\omega\sigma_1)$. If we adopt the typical values of

C, T_m, X_0 and U_0 worked out in (4.6), equation (87), and table 1, e.g.

$$C = 0{\cdot}9987 \times 10^{27} \text{ cm./sec}^3,$$
$$T_m = 2{\cdot}0154 \times 10^{-10} \text{ sec.},$$
$$X_0 = 113{\cdot}96 \times 10^8 \text{ e.m.u.},$$
$$U_0 = 2{\cdot}2330 \times 10^7 \text{ cm./sec.},$$

we obtain, using equations (117) and (119),

$$\sigma_2 = \frac{(1{\cdot}8254 \times 10^{-20}) - (2{\cdot}0154 \times 10^{-10}\, T) + 0{\cdot}5 T^2}{(2{\cdot}2360 \times 10^{-20}) - (2{\cdot}0154 \times 10^{-10}\, T) + 0{\cdot}5 T^2} \qquad \ldots\ldots(117a),$$

$$\underset{\omega \to 0}{\mathrm{Lt}}\,(\omega\sigma_1) = \frac{-2{\cdot}0127 \times 10^{17} + 0{\cdot}9987 \times 10^{27}\, T}{(2{\cdot}2330 \times 10^7) - (2{\cdot}0127 \times 10^{17} T) + (0{\cdot}4999 \times 10^{27} T^2)} \qquad \ldots\ldots(119a).$$

If we give T its value corresponding to transit time between cathode and anode, we have, taking $V = 300$ volts and $d = 0{\cdot}5$ cm. from table 1:

$$T = \frac{3d}{\sqrt{(-2Ve/m)}} = \frac{1{\cdot}5}{\sqrt{(3{\cdot}54 \times 3 \times 10^{17})}} = 1{\cdot}4157 \times 10^{-9} \text{ sec.}$$

We obtain further

$$\sigma_2 = \frac{1{\cdot}8254 \times 10^{-20} - 2{\cdot}8529 \times 10^{-19} + 1{\cdot}0020 \times 10^{-18}}{2{\cdot}2360 \times 10^{-20} - 2{\cdot}8529 \times 10^{-19} + 1{\cdot}0020 \times 10^{-18}} = 0{\cdot}9945,$$

$$\underset{\omega \to 0}{\mathrm{Lt}}\,(\omega\sigma_1) = \frac{-2{\cdot}0127 \times 10^{17} + 1{\cdot}4138 \times 10^{18}}{2{\cdot}2330 \times 10^7 - 2{\cdot}8490 \times 10^8 + 1{\cdot}0006 \times 10^9} = 1{\cdot}64 \times 10^9.$$

The above values are to be compared with the corresponding values obtained when a magnetic field of critical value is applied: under these conditions it is readily shown that at the anode

$$\sigma_2 = -1 + \frac{\omega^2 U_0}{C} + \frac{\left(\frac{e}{m} X_0\right)^2}{U_0 C} = 0{\cdot}940,$$

$$\omega\sigma_1 = \frac{eX_0}{mU_0} = -9{\cdot}02 \times 10^9,$$

which shows that the value of σ_2 at the anode is not greatly affected by a magnetic field, but that $\omega\sigma_1$ has a negative instead of a positive value. By comparison with equation (119a) we see that the value of $\omega\sigma_1$ at the anode in the critical condition is the same as its value at the cathode. But from equation (99)

$$\omega\sigma_1 = \frac{\omega}{U}\frac{dU}{d\theta} = \frac{1}{U}\frac{dU}{dT}.$$

In the critical condition U has the same value at the anode as at the cathode, namely U_0. Thus dU/dT also must have the same value in both places, namely $(e/m)\, X_0$. We see therefore that a virtual cathode exists just in front of the anode having just the same properties as the real cathode in respect of the forward component of direct-current velocity.

u_c (7.2) *Alternating velocity at critical plane.* When we come to consider the alternating velocity, we find, writing u_c for the value of u at the anode in the critical condition,

$$u_c = \frac{\beta}{p^2-\omega^2}\left[\left(1+\cos\frac{2\pi p}{\omega}+\frac{n_2}{1-n_1}\frac{\omega}{p}\sin\frac{2\pi p}{\omega}\right)\sin pt - \left\{\sin\frac{2\pi p}{\omega}+\frac{n_2}{1-n_1}\frac{\omega}{p}\left(1-\cos\frac{2\pi p}{\omega}\right)\right\}\cos pt\right] \quad \ldots\ldots(120),$$

in which
$$\frac{n_2}{1-n_1} = \frac{eX_0}{m\omega U_0} = -\frac{X_0}{U_0 H_c},$$

where H_c and X_0 are both positive.

Table 3 contains useful particular cases of the above expression corresponding to different values of p/ω.

Table 3

p/ω	u_c
0	$\frac{+\beta}{\omega^2}\left(\frac{2\pi X_0}{U_0 H_c}\right)\sin pt$
$\frac{1}{4}$	$\frac{16}{15}\frac{\beta}{\omega^2}\left[\left(1+\frac{4X_0}{U_0H_c}\right)\sin pt+\left(1-\frac{4X_0}{U_0H_c}\right)\cos pt\right]$
$\frac{1}{2}$	$\frac{8}{3}\frac{\beta}{\omega^2}\left[\sin pt-\frac{2X_0}{U_0H_c}\cos pt\right]$
$\frac{3}{4}$	$\frac{16}{7}\frac{\beta}{\omega^2}\left[\left(1-\frac{4X_0}{3U_0H_c}\right)\sin pt-\left(1+\frac{4X_0}{3U_0H_c}\right)\cos pt\right]$
1	$\frac{-\pi\beta}{\omega^2}\left[\cos pt+\frac{X_0}{U_0H_c}\sin pt\right]$
$\frac{5}{4}$	$\frac{16}{9}\frac{\beta}{\omega^2}\left[\left(\frac{-4X_0}{5U_0H_c}-1\right)\sin pt+\left(\frac{4X_0}{5U_0H_c}-1\right)\cos pt\right]$
$1\frac{1}{2}$	$\frac{8}{5}\frac{\beta}{\omega^2}\left[\frac{2X_0}{3U_0H_c}\cos pt-\sin pt\right]$
$\frac{7}{4}$	$\frac{16}{33}\frac{\beta}{\omega^2}\left[\left(1+\frac{4X_0}{7U_0H_c}\right)\cos pt-\left(1-\frac{4X_0}{7U_0H_c}\right)\sin pt\right]$
2	0

The alternating velocity u_c as given by equation (120) should strictly be regarded in the first instance as corresponding to the critical plane rather than to the anode. The critical condition corresponds, however, to the electrons coming to rest exactly at the anode, so that the distinction is apparent rather than real. The point was raised in order to prepare the way for a further conclusion which requires a little consideration. The question to be settled is as follows. Does the existence of an alternating component of velocity superimposed upon the steady component of velocity in any way alter the conditions under which equation (120) was derived?

In the first place we noted that whatever the value of u_c, the value of U remains unaffected. This may be shown by referring back to equations (44a) and (97).

Thus the value of θ is 2π during the whole cycle of events, θ/ω being the steady-state transit time. Thus we do not have to regard the critical plane as moving backwards and forwards in the ultradynamic condition, for the critical plane coincides with the anode during the whole cycle of events. This is, of course, in accordance with the original definition; but as it appeared in the section dealing only with the steady state, some further definition for the ultradynamic state might have been necessary. Actually we see that it is not, the position of the critical plane as defined in the steady state being preserved in the ultradynamic condition. We have thus achieved what we set out to do, namely to determine our ultradynamic solution at a fixed plane.

As to what actually happens to individual electrons in the critical condition, it is clear that we cannot investigate this point by reference to the solution for individual electrons as given at (*b*) in (2.42), since it is not permissible to take $\omega(t-t_0)$ as 2π to correspond with the critical condition.

Associated with the critical condition is the obvious conclusion that, during that half-cycle which corresponds to an anode potential below the mean value, electrons would, at low frequencies, fail to reach the anode at all, except for voltage-amplitudes comparable with the initial velocity U_0 expressed in volts. This conclusion is satisfactory in the case of negligible space charge, but the existence of space charge concentrated near the anode may alter the situation. To investigate this point, we may write down the electron current I_c at the critical plane as follows: I_c

$$I_c = P_c U_0 + \rho_c U_0 + u_c P_c + \rho_c u_c,$$

where P_c is the steady-state value of space-charge density at the critical plane, and P_c
ρ_c is the fluctuating component of space-charge density at the critical plane. Now ρ_c
electrons will reach the anode at all instants of the negative half cycle of u_c if I_c has the same sign as $(P+\rho_c)$, which requires that

$$U_0 > -u_c.$$

This shows that the conclusion still applies when space charges are present. This does not prevent u_c from being greater than U_0 during the positive half cycle, but equation (120) can then only be applied during the positive half cycle and that part of the negative half cycle for which the condition $|u_c/U_0| < 1$ is satisfied.

With the above reservations, figure 4 gives the alternating velocity u_c in arbitrary units for values of p/ω between 0 and 2, for the case where $X_0/U_0H_c = 4$, which corresponds closely enough to the values of X_0, U_0 and H_c used in the example standardized in previous sections and also for the case $X_0 = 0$ (dotted). In figure 4 is shown only the coefficient of $(-\sin pt)$, designated $(-\hat{u}_i)$. In figure 5 is given the value of $(-\hat{u}_i)$ corresponding to negligible space charge ($\theta_c = \pi$).

(7.3) *Alternating potential at critical plane.* Now the following equation connects the alternating potential v_c at the critical plane with the alternating velocity v_c
existing there:

$$\frac{-e}{m} v_c = U_0 u_c + \int_{\theta=0}^{\theta=2\pi} \frac{\partial u}{\partial t} \frac{dx}{d\theta} \, d\theta.$$

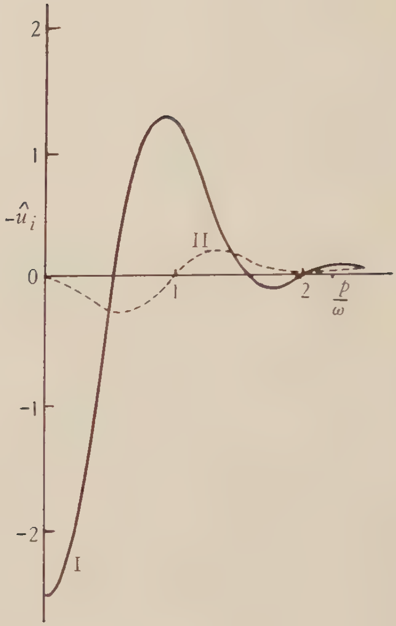


Figure 4.

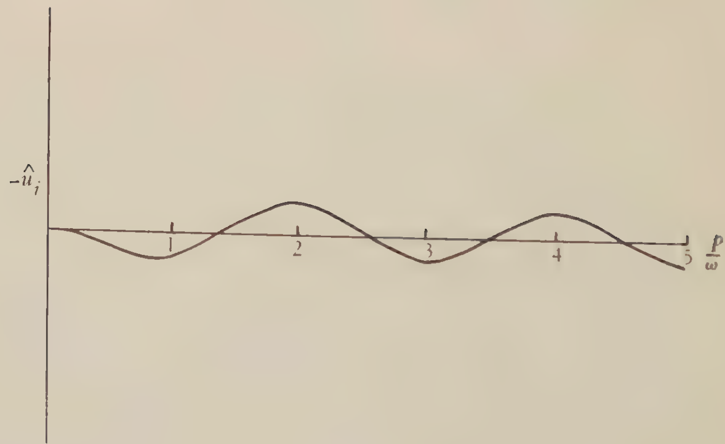


Figure 5.

Figure 4 may evidently be regarded as referring also to the alternating potential for values of p/ω which make the above integral of negligible value in comparison with $U_0 u_c$. The integral has not yet been evaluated, except in the case where U_0, X_0 are zero. In the absence of definite information regarding the importance of the integral term, figures 4 and 5 will be treated provisionally as referring to the alternating potential, subject to revision at some future date. The alternating potential will cut off in general during a part of the negative half-cycle, so that the dynamic condition has a distinct resemblance to the condition obtaining in an ordinary triode regenerative circuit. For $p/\omega = 1$ the peak value of v_c is

$$\frac{4\pi^2 m^2 c^2 X_0}{e^2 H_c^{\,3}} \dot{j}_1.$$

Taking X_0 as 114 V./cm., H_c as 131·6 G., $-j_1$ as 1 mA. or 10^{-4} e.m.u., we obtain $v_c = -0{\cdot}57$ V. The above condition corresponds to a negative conductance of the order of 2 mA./V.

(7.4) *Oscillatory properties of planar magnetron.* Inspection of figure 4 reveals that the coefficient of $(-\sin pt)$ has positive maxima when

$$p = \omega;\quad p = 2{\cdot}3\omega.$$

The relative insignificance of the maximum at $p = 2{\cdot}3\omega$ is in accordance with expectations (q.v.). The range of possible oscillation indicated by the positive values of $(-\sin pt)$ around the value $p/\omega = 1$ is as follows, when T_0 is written for $2\pi/p$

$$0{\cdot}625 < T_0/T < 1{\cdot}85 \qquad \text{......(121).}$$

The oscillations predicted by the range (equation 121) depend on the importance of electron-inertia and probably correspond to the characteristic oscillations of a solid-anode magnetron. The rapid clamping of the curve of figure 4 precludes the likelihood of obtaining any oscillations of higher frequency, the counterpart of which is well known in connection with the retarding field triode*. This result is in accordance with observation. On the other hand, figure 5 (for negligible space charge) *would* predict ranges of higher frequency oscillation. Oscillations with a cylindrical magnetron were produced by Zacek, originally in the year 1924†. An excellent summary of the present position of our experimental knowledge of magnetron oscillations is provided by Megaw‡. The corresponding oscillations in the retarded-field triode are the original Barkhausen-Kurz oscillations§. These have been extensively studied by workers too numerous to mention. Excellent summaries are to be found in papers by Hollman‖, Pierret¶, and Megaw**.

* Scheibe, *Ann. Phys.*, Lpz., **73**, 54 (1924); Potapenko, *Phys. Rev.* **39**, 625 (1932); Hollmann, *Z. f. Hochfrequenztechnik*, **37**, 145 (1931); Benham, *Electrical Communications XI*, **39**, 223 (1933).

† *Z. f. Hochfrequenztechnik*, **32**, 172 (1928).

‡ *Loc. cit.*

§ *Phys. Z.* **21**, 1 (1920).

‖ *Hochfrequenztech. u. Elektroakust.* **44**, 2, August 1934.

¶ *L'Onde Electrique*, **8**, 373 (1929).

** *J. Instn elect. Engrs*, **72**, 313 (1933).

(7.5) *The effect of cathode-temperature on oscillation-intensity.* The existence of an optimum in the relation between cathode-temperature and oscillation-intensity has been known since 1920*, but no satisfactory explanation has to my knowledge been proposed. The present paper contains all the material necessary for the establishment of a relation between cathode-temperature and the high-frequency energy developed by a planar magnetron diode in the critical condition. Comparison of theoretical results with experiment will constitute as severe a test as any of the fundamental correctness of the analysis.

In order to obtain the desired relation we shall first of all require to know the relation between cathode-temperature and total emission. It has been found that Richardson's law† holds quite accurately for any one coated cathode, provided it has been properly formed and aged. Measurements on several oxide-coated cathodes
b have led to the following average value of Richardson's constant b

$$b = 1{\cdot}42 \times 10^4\ {}^\circ\mathrm{K}.$$

For a total emission of 200 mA./cm.² at a value 1173° K. of T_c the corresponding value of a in the simplified Richardson equation $J_s = a\epsilon^{-b/T_c}$, with J_s in mA./cm.², is given by

$$\log_{10} a = 14{\cdot}4.$$

With the above information and with the assistance of (4.5) and (4.6) we obtain the results in table 4 for J_s, U_0 and X_0 as functions of T_c, the anode current being supposed constant at 50 mA./cm.²

Table 4

$J = 50$ mA./cm.²

T_c (° K.)	J_s (mA./cm.²)	$\frac{J_s}{J}$	$-V_m$ (V.)	$10^3\, x_m$ (cm.)	$10^{10}\, T_m$ (sec.)	X_0 (V./cm.)	$10^{-7}\, U_0$ (cm./sec.)	$10^7 X_0 U_0^{-1}$ (V.-sec./cm.²)
1279	2000	40	0·407	2·426	3·030	171·4	3·86	44·4
1235	800	16	0·2953	2·225	2·700	152·6	3·23	47·2
1222	600	12	0·2618	2·150	2·580	146	3·05	47·9
1190	300	6	0·1838	1·920	2·165	122·5	2·446	50·1
1173	200	4	0·1402	1·775	2·015	114·0	2·233	51·1
1145	100	2	0·0685	1·330	1·498	84·8	1·637	51·8
1132	75	1·5	0·0396	1·056	1·17	66·2	1·360	48·7
1125	60	1·2	0·0178	0·733	0·810	45·8	1·126	40·7
1118	50	1	0	0	0	0	—	0

The last column of table 4 gives the ratio X_0/U_0 which from table 3 is seen to constitute an important factor in determining the value of u_c. In particular, we note that when $p = \omega$ the coefficient of $(-\sin pt)$ is directly proportional to X_0/U_0. For a given value of β which is determined by j_1 we may say that the effect of cathode-temperature on the oscillation-intensity is to be measured by the ratio X_0/U_0. This ratio is seen from figure 6 to be a maximum when $T_c = 1145°$ K. In practice the optimum will occur below or above this temperature according to

* Barkhausen and Kurz, *loc. cit.*

† *Loc. cit.*

whether the emission-density when $T_c = 1173°$ K. is above or below the value here assumed, e.g. 200 mA./cm.²

The form of curve shown in figure 6 is quite characteristic of experimental behaviour, though in some classes of oscillation* a maximum at one temperature has been found to be followed by a minimum at some higher temperature. No trace of this minimum has been discovered theoretically by working with still higher emission-densities, but the reason for this is not difficult to see. Table 4 is adjusted to a fixed anode current of 50 mA., whereas, in practice, if the cathode-temperature be progressively raised the anode current will continually rise. This applies in the

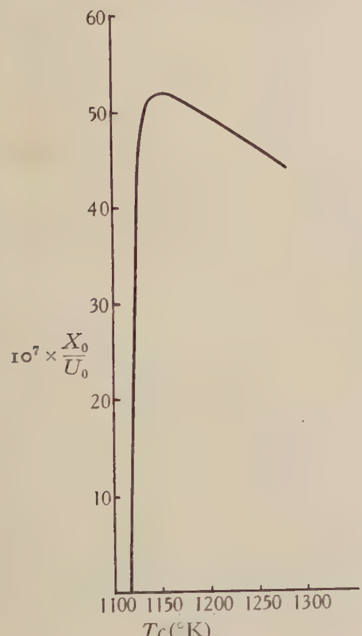


Figure 6.

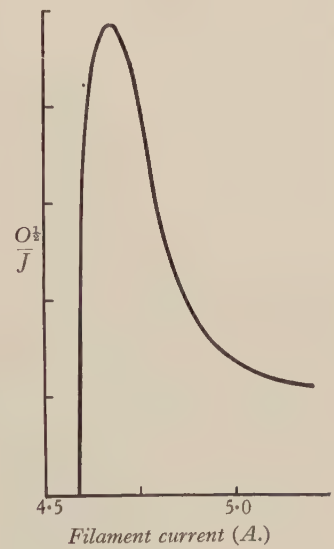


Figure 7.

case of coated cathodes for values of total emission up to many times the value of space current taken. This effect is well known among valve engineers as "poor saturation", though this term is also frequently applied to the other end of the scale (voltage saturation) in cases where the lack of a flat top to the anode characteristic is observed. Poor (space-charge) saturation is less noticeable with bright tungsten, but is still very noticeable when the tungsten filaments are external to the anode, as in McPetrie's case. The output curve as given by McPetrie corresponds to an anode current continually increasing as the temperature of the cathode is raised. A continually increasing value of J would be accompanied by a continually increasing value of j_1 which would *per se* tend to increase the oscillatory output as the temperature is raised.

* See McPetrie, *Wireless Engr*, p. 121 (1934).

Now McPetrie's curves show output as a function of filament current. The high-frequency voltage developed will be proportional to the square root of the output power O. If we divide this by the anode current we shall obtain a curve suitable for comparison with figure 6. The quantity $O^{\frac{1}{2}}/J$ is shown roughly on figure 7, and it is seen that there is no minimum following the maximum. In view of the marked difference in geometry and operating temperature and other differences, further information is required in order to establish quantitative agreement between theory and experiment.

In so far as theory predicts an optimum operating-temperature, the agreement with experiment may be regarded as promising. It will be seen that the theoretical optimum corresponds to an anode current of roughly half the saturation value, whereas experimentally the optimum is generally considered to occur for anode currents nearly equal to the saturation value. This difference may be due in part to the difference in geometry and operating-temperature but possibly also to a conservative estimate of the total emission. The accurate measurement of total emission requires a circuit which applies to the valve a sufficiently high voltage for a time insufficient to harm the cathode. The description of a suitable circuit for total-emission measurements is beyond the scope of the present paper, but it is of interest to note that emissions well over 2000 mA. cm.$^{-2}$ (see the first row of table 4) have been measured directly.

(7.6) *General conclusions concerning electron oscillations.* The expression (equation 120) for the alternating velocity existing at the critical plane of a magnetron is thought to be of fundamental importance as typifying the nature of virtual cathodes generally. Thus, whilst the virtual cathode as now understood is obtained by electrical and not by magnetic means, the only essential difference appears to lie in the existence in the magnetron case of a drift velocity parallel to the plates. It is possible that equation (120), suitably modified, may be used for the retarding-field triode. If T denotes the transit time between the real and virtual cathode of a retarding-field triode subject to no magnetic field the corresponding value of u_c is likely to be given by an expression of the type

$$u_c=\frac{\beta}{p^2}\left[\left(-1+\cos pT-\frac{eX_0}{mpU_0}\sin pT\right)\sin pt\right.$$
$$\left.-\left\{\sin pT+\frac{eX_0}{mpU_0}\frac{pT}{2\pi}(1-\cos pT)\right\}\cos pt\right] \quad \text{......(122)},$$

which differs from equation (120) only in that we have replaced $(p^2-\omega^2)$ by p^2 and $2\pi/\omega$ by T. Actually equation (122) as it stands does not remain finite when $p=0$.

L. Tonks* was the first to draw attention to the important part played by the formation of a virtual cathode in the retarding-field triode, and proposed some sort of combination of his theory for the steady state with the dynamic theory of Gill and Morrell†. The present paper is directed largely towards the fulfilment of the

* *Loc. cit.* † *Phil. Mag.* **44**, 161 (1922).

requirements postulated by Tonks, and it is believed that the analysis here given for the plane magnetron is accurate as far as it goes. The cylindrical case is likely to present grave difficulties judging by preliminary analysis, and it is to be hoped that many of the conclusions arrived at as a result of the study of a plane geometry will not be substantially modified when the results for cylinders are eventually derived.

§ 8. ACKNOWLEDGMENTS

In conclusion I should like to express my indebtedness to Marconi's Wireless Telegraph Company Limited for facilities in connection with the preparation of this paper, which is published by their permission.

APPENDIX. CHANGE OF COORDINATES

The theory as given in two earlier papers* involves the solution of the equation

$$\left(\frac{\partial}{\partial t}+U\frac{\partial}{\partial x}\right)\left(\frac{\partial U}{\partial t}+U\frac{\partial U}{\partial x}\right)=\frac{4\pi c^2 e}{m}J \qquad \text{......(1).}$$

Equation (1) is the electrical counterpart of Euler's form in which x and t are independent variables of the equation of motion for fluids moving in streamlines†. If we wish to employ coordinates which specify the position of an individual electron in the course of its passage between cathode and anode, we shall need to specify that the distance travelled by a chosen electron is dependent on the time (measured from some arbitrary zero) and also on the instant at which the electron in question started from the cathode. x, t

Taking t' and t_0 as Lagrangian variables, we shall have t', t_0

$$x=\phi\,(t', t_0),$$

$$t=t',$$

i.e. the electron which started at time t_0 (which $<t'$) has reached the plane x at time t'.

Let f be any function expressed in Eulerian coordinates, and $\bar{f}$ the same function when expressed in Lagrangian coordinates, i.e. $f, \bar{f}$

$$f(x, t)=\bar{f}(t', t_0).$$

Then
$$\frac{\partial f}{\partial t}\delta t+\frac{\partial f}{\partial x}\delta x=\frac{\partial \bar{f}}{\partial t'}\delta t'+\frac{\partial \bar{f}}{\partial t_0}\delta t_0,$$

or
$$\frac{\partial f}{\partial t}\delta t+\frac{\partial f}{\partial x}\left(\frac{\partial \phi}{\partial t'}\delta t'+\frac{\partial \phi}{\partial t_0}\delta t_0\right)=\frac{\partial \bar{f}}{\partial t'}\delta t'+\frac{\partial \bar{f}}{\partial t_0}\delta t_0,$$

and, since
$$\delta t'=\delta t, \quad U=\frac{\partial \phi}{\partial t'},$$

* Part I: *Phil. Mag.* 5, 641–62 (1928); Part II: *Phil. Mag.* 11, 457–517 (1931).
† See Lamb, *Hydrodynamics*, p. 2 (fifth edition).

we have, comparing coefficients of ∂t, ∂t_0 respectively,

$$\left.\begin{aligned} \frac{\partial f}{\partial t}+U\frac{\partial f}{\partial x}&=\frac{\partial \bar{f}}{\partial t'} \\ \frac{\partial \phi}{\partial t_0}\cdot\frac{\partial f}{\partial x}&=\frac{\partial f}{\partial t_0} \end{aligned}\right\} \qquad \text{......(2).}$$

Equations (2) are the formulae for change of variable, and show that, in the Lagrangian scheme, equation (1) simplifies, if we now write $t'=t$, to

$$\frac{\partial^3 x}{\partial t^3}=\frac{4\pi c^2 e}{m}J,$$

in which the independent variables are t and t_0 though the latter does not appear explicitly.

Writing $$\frac{4\pi c^2 e}{m}J=B+\beta\sin pt,$$

and integrating with respect to t, while t_0 is constant, we have (assuming the electron starts with zero velocity and acceleration)

$$\frac{\partial^2 x}{\partial t^2}=B\,(t-t_0)+\frac{\beta}{p}(\cos pt_0-\cos pt),$$

$$U=\frac{\partial x}{\partial t}=\frac{B}{2}(t-t_0)^2+\chi \qquad \text{......(3),}$$

$$x=\frac{B}{6}(t-t_0)^3+\psi \qquad \text{......(4),}$$

in which we have abbreviated by writing

χ $$\chi=\frac{b}{p}(t-t_0)\cos pt_0+\frac{b}{p^2}(\sin pt_0-\sin pt),$$

ψ and $$\psi=\frac{b}{2p^3}[p^2(t-t_0)^2\cos pt_0+2p\,(t-t_0)\sin pt_0+2\,(\cos pt-\cos pt_0)].$$

Having obtained a solution for U (equal to $\partial x/\partial t$) which is very simply arrived at, we seek to express the solution in the original coordinates in order to enable us to calculate physically interesting quantities, such as conductance, admittance, etc. From equation (3) we obtain

$$(t-t_0)=\left[\frac{6\,(x-\psi)}{B}\right]^{\frac{1}{3}},$$

$$\therefore\ \frac{B}{2}(t-t_0)^2=\frac{B}{2}\left(\frac{6}{B}\right)^{\frac{2}{3}}(x-\psi)^{\frac{2}{3}}$$

$$=U_1\left(1-\frac{\psi}{x}\right)^{\frac{2}{3}},$$

where we have written U_1 for the value of $\partial x/\partial t$ in the steady state ($\psi=0$).

δT Writing U instead of $\partial x/\partial t$, and $(T+\delta T)$ instead of $(t-t_0)$, where δT is to be regarded as the deviation from the mean or steady state value of transit time, we

obtain, from (2) and (4),

$$U = U_1\left(1-\frac{\psi}{x}\right)^{\frac{2}{3}} + \bar{\chi}\,(t', t_0)$$
$$= U_1\left(1-\frac{\psi}{x}\right)^{\frac{2}{3}} + \chi\,(T, t) + \text{terms in } \delta T.$$

It is readily shown that the fundamental component of U contains no term in δT, while the sole contribution to the fundamental arising from the term in ψ is

$$-\frac{2U_1}{3x}\psi.$$

Writing $$U_1 x^{-1} = 3T^{-1},$$

we obtain, for the fundamental u,

$$u = -2T^{-1}\psi + \chi.$$

Giving χ and ψ their values, we have

$$u = \frac{\beta}{p^2}\left[-\sin pt - \sin pt_0 + \frac{2}{pT}\cos pt_0 - \frac{2}{pT}\sin pt_0\right],$$

in which $$t_0 \equiv t - T,$$

in agreement with the solution given on p. 648 of part I.

Let us now derive the potential in Lagrangian coordinates. In Eulerian coordinates we have

$$-\frac{e}{m}\frac{\partial V}{\partial x} = \frac{\partial U}{\partial t} + U\frac{\partial U}{\partial x}.$$

In Lagrangian coordinates t, t_0 we have, using equations (2),

$$-\frac{e}{m}\frac{\partial V}{\partial t_0} = \frac{\partial x}{\partial t_0}\cdot\frac{\partial^2 x}{\partial t^2};$$

$\partial x/\partial t_0$ may be obtained from (3), thus

$$\frac{\partial x}{\partial t_0} = -\frac{B}{2}(t-t_0)^2(B+\beta\sin pt_0).$$

Hence

$$\frac{\partial x}{\partial t_0}\cdot\frac{\partial^2 x}{\partial t^2} = -\frac{B^2}{2}(t-t_0)^3 - \frac{B\beta}{2}(t-t_0)^3\sin pt_0 - \frac{B\beta}{2}\frac{(t-t_0)^2}{p}(\cos pt_0 - \cos pt)$$
$$-\frac{B^2}{2}\frac{(t-t_0)^2}{p}(\sin pt_0\cos pt_0 - \sin pt_0\cos pt).$$

This equation shows incidentally that harmonics are present in the potential if the current through the system is free from harmonics. Interesting ourselves only with the fundamental, we shall omit the last term. Integration with respect to t_0 yields the result

$$\frac{-e}{m}V = \frac{B^2T^4}{8} + \frac{B\beta}{2p^4}[p^3T^3\cos pt_0 + 2p^2T^2\sin pt_0 - 4pT\cos pt_0$$
$$-\tfrac{1}{3}p^3T^3\cos pt + 4(\sin pt - \sin pt_0)],$$

where $T \equiv t - t_0$.

The above equation may be reconverted to Eulerian coordinates by a process similar to that adopted in the case of the velocity, and the result agrees with the solution obtained by working in Eulerian coordinates throughout (see p. 510 of part II).

Turning now to the cylindrical case, equations (41) and (42) on p. 501 of part II,

$$\frac{\partial U}{\partial t}+U\frac{\partial U}{\partial r}=\frac{-e}{m}\frac{\partial V}{\partial r},$$

$$\left(\frac{\partial}{\partial t}+U\frac{\partial}{\partial r}\right)\left[r\left(\frac{\partial U}{\partial t}+U\frac{\partial U}{\partial r}\right)\right]=\frac{2ec^2}{m}J.$$

These become, as before,

$$\frac{\partial r}{\partial t_0}\cdot\frac{\partial^2 r}{\partial t^2}=\frac{-e}{m}\frac{\partial V}{\partial t_0},$$

$$\frac{\partial}{\partial t}\left(r\frac{\partial^2 r}{\partial t^2}\right)=\frac{2ec^2}{m}J.$$

The last equation requires the integration of an equation of the type

$$r\frac{\partial^2 r}{\partial t^2}=f(t).$$

The difficulties in connection with such an integration are such as to render a treatment in Lagrangian coordinates hardly less intractable than the treatment already given in Eulerian coordinates on p. 501 of part II.

Summarizing the above, we see that by means of a suitable change of variable it is possible to avoid the necessity for solving tiresome differential equations. A simplification of the equations results also in the cylindrical case, which, however, does not appear to permit of exact solution in either system of coordinates. The above conclusions break down in more complicated cases. Thus if initial velocities and accelerations, or a magnetic field, be included, the process becomes intractable and the method may become too inaccurate to be of value.

DISCUSSION

Prof. C. L. Fortescue. What is the exact mechanism whereby a two-electrode valve can have amplifying properties which give it the nature of a negative resistance? And are there any experiments to prove the author's theory to be correct—can a parallel-plate diode oscillate? Llewellyn says that this is probably impossible.

Author's reply. I find Prof. Fortescue's first question a little difficult to understand. The negative-resistance property is hardly less fundamental than the electron inertia, which is responsible for phase-differences corresponding to a negative power factor under favourable conditions. The amplifying properties, if any, would in my opinion be incidental to the negative resistance, and depend also on the external circuit. The mechanism of electronic oscillations in diodes and triodes is

dealt with in an article due to appear in the *Wireless Engineer*. With regard to experimental confirmation, Müller* has recently established the existence of oscillations in a carefully constructed low-loss full-wave parallel-plane diode. These oscillations lie within the zone predicted by my theory, which also predicts the correct intervals for dwarf waves in a triode oscillator†.

The above remarks all refer to diodes or triodes without magnetic field. So far oscillations in a parallel-plane diode with magnetic field have not to my knowledge been observed. Any experiments conducted with a view to obtaining such oscillations would be subject to the difficulty of obtaining a sufficiently uniform magnetic field throughout the space between the plane electrodes. If these electrodes are not large compared with their separation edge, effects occur which may mask the oscillations or which may result in spurious oscillations due to a proportion of electrons missing the anode and returning to it from behind.

* *Hochfrequenztech. u. Electroakust.* **43**, 6, 195–199 (1934).
† *Elec. Comm. XI*, **39**, 223 (1933).

537·743

THE ABSOLUTE MEASUREMENT OF ELECTRICAL RESISTANCE BY A NEW ROTATING-COIL METHOD

By H. R. NETTLETON, D.Sc., Lecturer in Physics, Birkbeck College
AND E. G. BALLS, M.C., B.Sc., A.I.C., Birkbeck College

Received July 11, 1934. *Read November* 2, 1934.

ABSTRACT. A rotating coil of mean radius α lies symmetrically between two fixed twin coils of radius a. If the ratio α/a lies between 0·58 and 0·53 it is easy to arrange, by merely adjusting the distance between the twin coils, that the mutual inductance M between the rotating coil and the two fixed coils shall be very accurately proportional to the angle θ of displacement from the conjugate positions over a range of some 10° of arc on either side of the zeros. The constant K of the relation $M=K\theta$ can then be accurately measured by a method here described in which θ is deduced from a mutual-inductance ratio. At the same time errors in the calibration of the inductometer used are eliminated.

If such a coil spins with an angular velocity ω, while a current C traverses the fixed twin coils, a uniform e.m.f. ωCK can be drawn off a commutator on the rotating shaft and made to balance an e.m.f. CR drawn off an adjustable resistance R carrying the same current. Thus $R=\omega K$.

If now the twin coils are brought rather closer together, the law of variation of M with θ takes the form $M/\theta=K+A\theta^2-B\theta^4$ over a displacement of 35° of arc from the zeros, where A and B are very small positive constants and M/θ attains a maximum value K', where $\theta^2=A/2B$ and scarcely changes over a range of 3° in this neighbourhood. This allows larger sectors to be used on the commutator and the constant K' of the relationship $R=K'\omega$ can be accurately determined for a suitable commutator *in situ*.

The experimental work is mainly devoted to a study of the laws of inductance on which the method depends and to the determination of the constants K and K', but preliminary spin experiments are very hopeful and resistances of between 0·32 Ω. and 0·64 Ω. have been measured absolutely by means of commutators with sector contacts of 23° and 47° of arc. Owing to the relatively large e.m.f.'s involved, the method is very sensitive and a fluxmeter can be used as the balance-detector for hand-controlled stroboscopic spins. In other experiments a synchronized television motor was used.

§ 1. INTRODUCTION

The well-known method of measuring electrical resistance absolutely with the aid of a spinning coil was originally suggested by Weber and put forward independently by Kelvin to the Electrical Standards Committee of the British Association in 1863. It is very fully described in the reports of the British Association covering the period 1862–67; the experiments were carried out principally by Maxwell, Stewart, and Jenkin. Later determinations by this method were undertaken by Rayleigh and Schuster*; Rayleigh† and H. Weber‡.

* *Proc. roy. Soc.* **32**, 104 (1881). † *Phil. Trans.* **173**, 661 (1882).
‡ *Der Rotations-inductor* (Leipzig, Teubner, 1882).

The method involves fundamentally the measurement of a coil-area, a galvanometer constant, a speed of rotation, and the deflection of a needle at the centre of the spinning coil. Corrections are necessary for the moment of the magnetic needle, the torsion of the supporting fibre, and the self inductance of the coil. It is not surprising, therefore, that the method has largely given way to the methods of Lorentz and Campbell which involve quantities more easy to determine and are applicable to the measurement of external resistances with potential-leads.

In 1880 Carey Foster* suggested an interesting null method involving the same principle as the British Association method. In this arrangement a steady current is passed through a tangent galvanometer of known principal constant and through the resistance to be measured. The e.m.f. across the potential leads of this resistance is balanced against that derived from a coil spinning in the earth's field, the balancing circuit being completed through commutators only over some 20° of arc. The middle of the period of contact was made to coincide with the instant when maximum e.m.f. was induced in the spinning coil, and the extreme variations of the e.m.f. during contact was 1·83 per cent. Though this null method has the great advantage of dispensing with the corrections necessary in the original method and is applicable to the measurement of an external resistance, the same fundamental quantities are involved. The correction necessary for the angle of contact, which must be measured, depends for its validity on a symmetrical setting of the commutators about the position of maximum e.m.f. which is difficult to locate with precision, and thermoelectric effects are likely to be troublesome.

The object of the present communication is to describe a preliminary investigation of a sensitive null method by which a resistance may be measured absolutely in terms of a mutual inductance and a frequency.

§ 2. THE THEORY OF THE METHOD

Simple form of the method. A coil spinning uniformly about a horizontal diameter lies between two larger fixed twin field coils having their planes horizontal and so separated in the theoretically simplest type of experiment that the mutual inductance M between the rotating and fixed coils in series and conjunction is, over a range of some 10° of arc on either side of zero, very accurately proportional to the angle θ of displacement from the position of zero mutual inductance. The axis of rotation lies in the magnetic meridian and the earth's vertical flux through the rotating coil is neutralized by a small current passing through large compensating coils in the Helmholtz position.

A steady current C of about one ampere is passed through the twin fixed coils and through a variable manganin resistance which is adjusted until the e.m.f. across its potential-leads is balanced by that across the commutating sectors of the rotating coil. These make contact through fixed brushes with a galvanometer over some 20° of arc, during which the e.m.f. arising from the uniform spin is constant.

* *B.A. Reports*, p. 426 (1881).

M Since the mutual inductance M at any position θ over the range of contact is given by the linear relationship

$$M=K\theta \qquad \text{......(1)},$$

K where K is constant, we have for the flux F through the spinning coil at any position θ

C
$$F=CM=CK\theta \qquad \text{......(2)},$$

and for the numerical value of the e.m.f.

E, t
$$E=dF/dt=CK\omega \qquad \text{......(3)},$$

ω where ω is the constant angular velocity of rotation. Whence, equating this to the
R e.m.f. CR across the balancing resistance R, we have

$$R=K\omega=2\pi nK \qquad \text{......(4)},$$

n where n is the number of revolutions per second.

In view of the remarkable accuracy of the linear law connecting M and θ over considerable range the constant K can be determined with precision, and a method of doing this is described below which not only avoids the direct measurement of angles but at the same time eliminates errors in the stud-calibration of the mutual inductometer used. Further, on account of the linear law which gives rise to a uniform e.m.f. the exact angle of contact is not required and, although symmetry of contact is readily obtainable by means of inductance measurements, slight departure from symmetry is unimportant. Thermoelectric effects are rendered insignificant by adjusting the earth-coil current for zero galvanometer deflection when the current C is broken and the resistance R is connected across the contacts of the rotating coil spinning at the frequency n. On application of the current C relatively large opposing e.m.f's. come into action, giving high sensitivity, and on reversal of C throughout the balance is preserved.

Extension of the method to large angles of contact. So far we have supposed that the twin field coils are so separated that the linear law $M=K\theta$ is rigorous and that an unvarying e.m.f. $CK\omega$ is drawn off the rotating coil over an angle limited to some 20° of arc. Let us now consider a more general case applicable when the commutator sectors are enlarged to some 50° of arc.

M' θ' M'' θ'' Let M' be the change in the mutual inductance between the rotating coil and the twin fixed coils over the angle of contact θ' of the sectors with the brushes in the region of the first conjugate position, and let M'' be the change in mutual inductance over the corresponding angle of contact θ'' around the second conjugate position. Then the flux changes CM', CM'' occur in times θ'/ω, θ''/ω respectively. Thus the average e.m.f. during the contacts of one revolution is $C\omega\,(M'/\theta'+M''/\theta'')/2$; and if this is balanced on a ballistic galvanometer or fluxmeter by the opposing e.m.f. CR, drawn off the adjustable resistance R, we have:

$$R=\omega\frac{1}{2}\left(\frac{M'}{\theta'}+\frac{M''}{\theta''}\right)=\omega K' \qquad \text{......(5)}.$$

Now while in general it would not be possible to measure precisely the sweeps of inductance M' and M'' over precise sector angles θ' and θ'', it is easily possible

to measure with great precision the ratios M'/θ' and M''/θ'' provided that M/θ is rendered constant in the neighbourhoods of the sector-extremities. This, it has been found, can be readily accomplished by bringing the twin field coils rather closer together than when they are set to yield the linear law, so that their mutual inductance for a range of some 35° of arc on both sides of the zero obeys a law of the form

$$M/\theta = K + A\theta^2 - B\theta^4 \qquad \text{......(6).}$$

It will be seen below that the coils were so separated that

$$M/\theta = 124{\cdot}932 + 3{\cdot}665 \times 10^{-4}\theta^2 - 3{\cdot}05 \times 10^{-7}\theta^4 \qquad \text{......(7),}$$

where M is in nominal microhenries and θ is in degrees of arc. This gives a maximum value for M/θ of 125·042 μH. per degree at 24°·5. The value 125·040 corresponds to both 22°·8 and 26°·1, so that over this range of 3°·3 of arc M/θ may be regarded as constant and equal to $dM/d\theta$.

By making one sector larger than the other so that the range of contact is determined by the smaller sector, θ' and θ'' were made equal at a value of about 46°·8. M'/θ' and M''/θ'', rendered almost equal by symmetrical setting, were measured by observations of M and θ in the neighbourhoods of the sector edges, readings being taken when the sectors were (*a*) just on and (*b*) just off the contact brushes, the differences in angle between the on and the off readings being narrowed down to about 0°·1. Under such circumstances, the constant K' can be measured with high precision.

It should be observed that although in this method of working with a large angle of contact the e.m.f. drawn off the rotating coil is not quite constant, the extreme variation of e.m.f. is only 0·16 per cent over a contact of 47° as against 1·83 per cent over 22° of arc in Carey Foster's method. Moreover in this method the e.m.f. at the point of leaving the sectors is equal to the average e.m.f. over the whole contact.

Design and arrangement of coils. The design and arrangement of coils which provide the simple inductance laws stated above are based on the following theory. If two concentric circles have the ratio of their radii α/a equal to 0·506078* the mutual inductance between them is so accurately proportional to the angle of displacement from the conjugate positions that the rising deviation from a straight-line law amounts to only 4·2 parts in a million at 7° of displacement and 18 parts in a million at 10° of displacement. With a slightly larger ratio of α/a, the deviation may be distributed so that over a range of 7° it never exceeds 0·75 parts in a million. On the other hand, if a coil of radius α lies between two circles of radius a separated by a distance $2x$, the ratio α/a must be increased to bring about a similar approach to linearity. Thus if $x/a = 0{\cdot}19438$, the ratio α/a must be raised to 0·5461 to preserve the limiting linear law, the deviation at 7° being now 5·1 parts in a million. This deviation may likewise be distributed and diminished by slightly raising the ratio α/a.

The practical significance of this theory, coupled with the fact that the primary effect of multiplicity of layers is to alter slightly the effective radii of the coils, lies

* Nettleton and Llewellyn, *Proc. phys. Soc.* **44**, 195 (1932).

in the result that if α/a lies between the limits $0{\cdot}58 > \alpha/a > 0{\cdot}52$, linearity within the accuracy of experimental measurements may be secured over some 12° with multiple-layered coils by merely adjusting the distance of separation between the larger twin coils. Such twin coils may at any time be joined in opposition, thus enabling the smaller coil to be set symmetrically.

In general we may express the mutual inductance between twin coils and a smaller coil displaced from the conjugate positions by an angle θ by a series of the type

$$M = K\theta + A\theta^3 + B\theta^5 + \ldots\ldots \qquad \ldots\ldots(8),$$

where the constants K, A, B etc. depend on α, a, x and the number of turns, and may be evaluated with the aid of Legendre functions, though the process is laborious.

If α/a lies between the limits given, the coefficient A may be rendered zero by adjusting the separation $2x$. The succeeding coefficients are then small and negative, and a limiting linear law with accuracy of the order already stated results. If now the separation $2x$ is reduced, A assumes a small positive value and we have in practice with high accuracy over 30° of arc

$$M = K\theta + A\theta^3 - B\theta^5 \qquad \ldots\ldots(9),$$

which gives a maximum value of $M/\theta = K + A^2/4B$ at positions given by $\theta^2 = A/2B$. These positions are those proposed for the sector edges and define the ideal angle of contact for the new and closer distance of separation.

The e.m.f. on uniform rotation is everywhere proportional to $dM/d\theta$. Thus the minimum e.m.f. over the sector contacts is represented by K when $\theta = 0$, the maximum e.m.f. by $K + 9A^2/20B$ when $\theta^2 = 3A/10B$, and the average e.m.f. by $K + A^2/4B$ (which is the actual e.m.f. when $\theta^2 = A/10B$ or at the sector edges) where $\theta^2 = A/2B$.

Measurement of angle by a mutual inductance method. This method of measuring an angle of displacement is particularly suitable for the present purpose and is based upon the fact that the mutual inductance between a small solenoid of designed dimensions and two large twin coils between which it rotates can be rendered with great accuracy proportional to the sine of the angle of displacement from the conjugate positions.

The mutual inductance between twin circles A and B, figure 1, and a single
L layered solenoid C of radius α and length $2L$, rotated from the conjugate position by an angle θ, is given by the expression

M_θ

$$M_\theta = Gq \sum_{n=1}^{n=\infty} \frac{2}{n(n+1)(n+2)} \left(\frac{s}{r}\right)^{n-1} P_n'(\cos\psi) \frac{P_{n+1}(\cos\phi)}{\cos\phi} P_n(\sin\theta) \ldots\ldots(10),$$

P_n, P_n' where P_n is the Legendre function of the first kind, of order n; P_n' its differential
G coefficient; n an odd positive integer, G the galvanometer constant of A and B
s, r, ϕ, ψ together at the origin of symmetry; q the total area of C; and s, r, ϕ, and ψ are sufficiently defined by the figure.

If the angle solenoid C is of small radius, this series converges rapidly and only

the first three terms are of importance. They lead to the expression

$$M_\theta/Gq = \sin\theta + K_1 P_3(\sin\theta) + K_2 P_5(\sin\theta) \qquad \ldots\ldots(11),$$

where

$$K_1 = \frac{3\cdot 4}{2}\cdot\frac{1}{r^4}\left[x^2 - \frac{a^2}{4}\right]\left[\frac{L^2}{3} - \frac{\alpha^2}{4}\right]$$

and

$$K_2 = \frac{5\cdot 6}{2}\cdot\frac{1}{r^8}\left[x^4 - \frac{3}{2}x^2a^2 + \frac{a^4}{8}\right]\left[\frac{L^4}{5} - \frac{L^2\alpha^2}{2} + \frac{\alpha^4}{8}\right],$$

or when the result is expressed in powers of $\sin\theta$

$$M_\theta/Gq = \sin\theta\,[1 - 3K_1/2 + 15K_2/8] + \sin^3\theta\,[5K_1/2 - 35K_2/4] + \sin^5\theta\,[63K_2/8] \qquad \ldots\ldots(12).$$

If α/a is less than $\frac{1}{4}$, and K_1 is rendered sensibly zero by making $L^2/3$ equal to $\alpha^2/4$ or x^2 equal to $a^2/4$, K_2 also is always small and a close approach to the sine law results. If the coils are multiple-layered to the extent used in this research, this effect may be treated by a method due to Maxwell* and shown to be negligibly small.

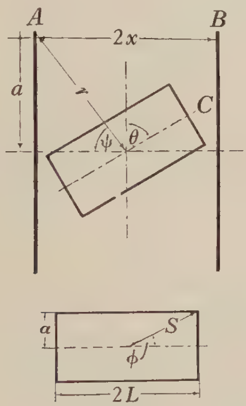


Figure 1.

In the angle solenoid used in our experiments $2L = 3{\cdot}78$ cm. and $\alpha = 2{\cdot}18$ cm. as found with the aid of a standard solenoid. The mean radius a of the twin coils was $16{\cdot}41$ cm. and in the least favorable position used, namely when the coils were closest and $\psi = 78°$, these figures give $K_1 = -1{\cdot}09\times10^{-5}$ and $K_2 = -3{\cdot}20\times10^{-5}$. Whence if $M_{\max.}$ denotes the value of M when $\theta = 90°$

$$\sin\theta = \frac{M_\theta}{M_{\max.}}\,[1 + 7\times10^{-7} - 2{\cdot}5\times10^{-4}(\sin^2\theta - \sin^4\theta)] \qquad \ldots\ldots(13)$$

and the sine law is very accurate, but it should be observed that the value of K_1 is necessarily uncertain as errors in the measurements of L and α have large effects on its calculated value. Some experimental tests upon this angle coil have already been described by Llewellyn†, but in view of the special advantage of the method for the present purpose and its sensitivity to two or three seconds of arc over considerable

* *Proc. phys. Soc.* **42**, 507 (1930).
† Thesis, Ph.D. degree, University of London.

range, further improvements in the method and in the tests of its accuracy are contemplated.

Method of eliminating local inductometer errors. Essentially the methods here described with either large or small angles of contact require the determination of a constant M/θ, the angle θ being measured from the relationship $\sin\theta = M_\theta/M_{\max}$. All mutual inductances are measured on a Campbell inductometer. This inductometer is first carefully calibrated so that all the stud readings are known in terms of 100 divisions of the scale, so that inductances can be read in nominal microhenries, subject to calibration errors. Further, the rotating coil is wound with such a number of turns that its mutual inductance with the twin field coils is so close to the value of the mutual inductance between the angle coil and the same field coils that up to at least 25° of arc the readings are within 100 microhenries; thus the same thousands and hundreds studs are in use for both readings. With the widening gap between the radian law and the sine law, the difference of reading at 40° of arc is still only some 300 μH.

Under such circumstances it is easy to show that any errors of the order possible in the calibration are rendered quite negligible as regards the determination of the ratio M/M_θ, and the accuracy of M/θ is solely dependent on the accuracy of the inductometer at the angle-coil reading $M_{\max.}$, which in our case was some 7000 μH.

In future work we shall aim at a maximum angle-coil reading of just over 10,000 μH., which will enable the fundamental dimensional length measurement to be checked by a 10 millihenry standard without any resort to the calibration curve.

§ 3. THE APPARATUS

The formers of the fixed twin field coils A and B, figure 2, were constructed of dexonite. The channels, of radial depth and axial breadth 3·6 cm., were each filled with 504 turns of double-silk-covered copper wire, s.w.g. 16, the mean diameter of winding being 32·8 cm. The coils, separated by distance pieces D, D, were locked together and firmly supported.

The rotating coil C consisted of 80 turns of double-silk-covered copper wire, s.w.g. 26, wound on a solid mahogany former, the channel having axial breadth 1 cm. and radial depth 0·3 cm. The mean diameter of winding was about 19·0 cm. A hole of diameter 5 cm. through the centre of the former served for the insertion of the angle solenoid when required. The hole was lengthened by attaching to each face of the former wooden rings, one of which is seen at E; to the ends of these rings were fastened small brass brackets with screw adjustments enabling the angle coil to be set centrally and secured in position.

The portions of the brass shaft to which the rotating coil C was attached terminated in pieces of channel brass F, F which held the coil by brass bolts. Some play within the channel pieces, which allowed room for packing, enabled the coil to be set symmetrically about the axis of rotation before the bolts were screwed up tightly. The brass shaft was of external diameter 1·27 cm. and had a central hole of bore 0·5 cm. which permitted bell flex leads from the rotating coil to be led through the portion G of the shaft to terminals t, t on the sectors of the commutator M.

The commutator M consisted of a cylindrical ebonite piece some 6·5 cm. in diameter and 3·3 cm. long, through the centre of which passed stout brass tubing which enabled it to be slipped over the shaft and fixed thereon by heavy screws. The brass sectors let into the ebonite were diametrically opposite. In the particular commutator shown inset in figure 2 on a larger scale, one sector S consisted of a brass piece tapering along its length from 44° to 51° of arc while the other sector was parallel along its length and of width equivalent to 60° of arc. The framework N carried spring brass brushes adjustable in width and height and provided with terminals. The fly-wheel L, the stroboscopic disc R, the flexible unions U and the well-separated motor T are seen in the diagram.

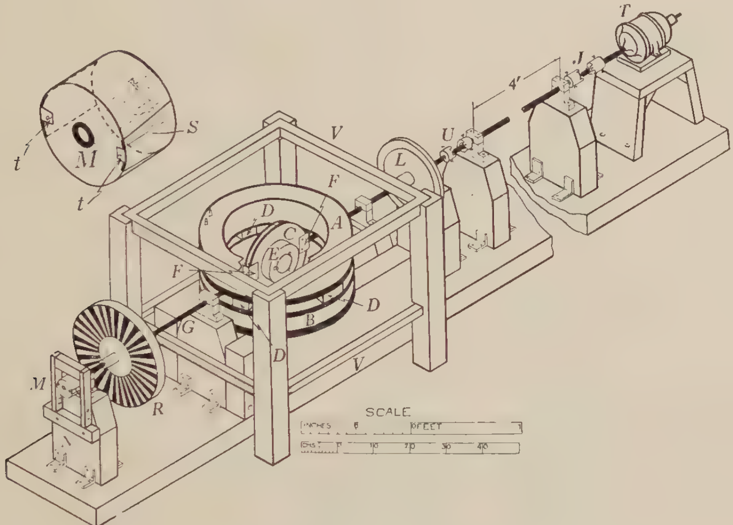


Figure 2.

The large horizontal coils V, V of sides 49×55 cm. are placed in the equivalent Helmholtz position and serve for the neutralization of the earth's vertical flux through the rotating coil when fed by a small current derived from accumulators.

The main electrical circuit will be readily understood from figure 3 with little description. The current of about 1 A., derived from accumulators, can be sent in either direction through the twin coils A and B, the quadrant key Q_3 serving to join them either in conjunction or in opposition but always in series. The same current passes through standardized resistances of 0·20029 Ω. and 0·50006 Ω. and through the variable resistance R to be adjusted and measured, all of which can be connected by potential leads with a good thermoelectric potentiometer which reads from 0 to 90 millivolts so that the value of R can be checked by comparison within a few minutes of measuring its resistance absolutely. The resistance R consisted of a box of manganin resistances of nominal values varying from 0·005 Ω. to 2·0 Ω. in series with a short semi-circular copper wire, of s.w.g. 18 provided with a movable potential contact. The e.m.f. across R can be adjusted to neutralize that across the

brushes b of the uniformly spinning coil, balance being observed on a galvanometer or fluxmeter provided with a tapping key.

The circuit also readily permits of the measurement of the mutual inductance between the rotating coil, when stationary, and the twin field coils as well as of the measurement of the angle between them. For this purpose the primary of a Campbell mutual inductometer is switched in at the quadrant key Q_1 to be in series with the fixed coils. The rotating coil through the sector terminals t, t is thrown into series with the secondary of the mutual inductometer at the quadrant key Q_2 and the galvanometer is included in this loop and detached from R by rocking over the mercury switch D. Alternatively, the angle coil may be thrown into the secondary circuit at H instead of the rotating coil and mutual inductance, and hence the angle between it and the twin coils can be measured. Balance is easily obtained to within 0·1 μH., which represents in the case of the angle coil 3″ of arc.

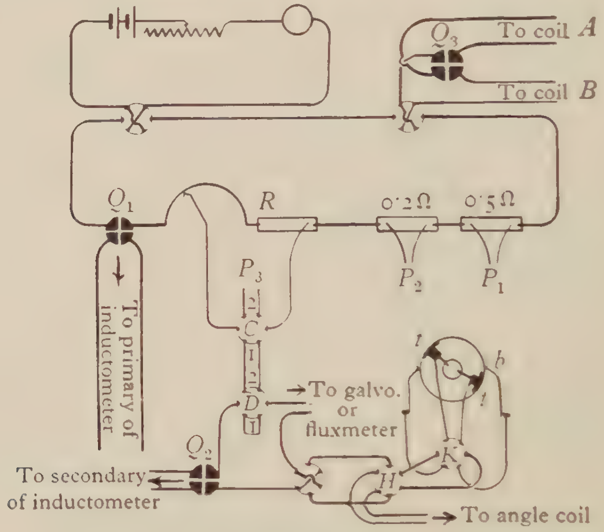


Figure 3.

The most constant speed of revolution was obtained by using a synchronized television motor, the cogged wheel of which had thirty narrow teeth separated by gaps four times the width of a tooth. The synchronizing impulse is fed to coils actuating an electromagnet pulling upon the teeth of the cogged wheel. The impulses used were derived from valve-maintained König tuning-forks which had been calibrated against one another by the method of visible beating*. In figure 4 the complete valve circuit is shown for maintaining the fork, synchronizing the motor and supplying a neon lamp with the tuning-fork frequency for viewing the stroboscopic disc R, figure 2, now provided with thirty lines. Because the television motor was rather weak in power, it was somewhat overrun and required careful rheostat adjustment to maintain synchronization for brief periods, but perfect

* Nettleton and Balls, *Proc. phys. Soc.* **45**, 545 (1933).

balances were obtained on the galvanometer at three different speeds, viz. 8·533, 10·667 and 12·800 revolutions per second corresponding respectively to forks of frequencies 256, 320, and 384 vibrations per second.

For other speeds a stronger motor was used in conjunction with other stroboscopic disc rulings, viewed by means of the fork-controlled neon lamp, and the average speed was maintained as constant as possible with rheostats and hand-friction control. Though the imperfections of this control were manifested on a galvanometer by the oscillations of the spot of light, excellent balance was obtained by means of the following artifice. A Grassot fluxmeter of the silk-fibre-suspension

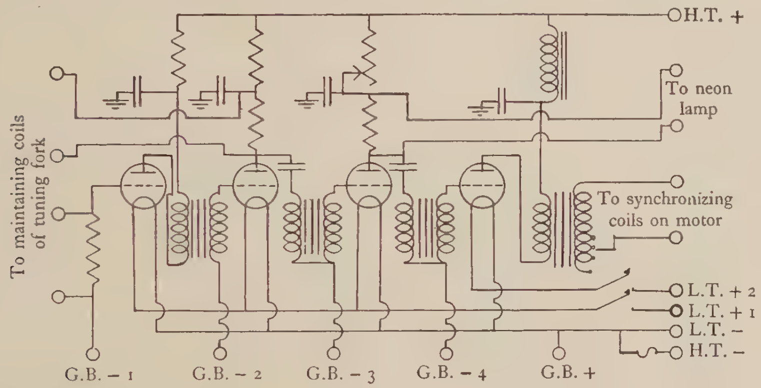


Figure 4.

type* shunted by some 80 Ω. was substituted by a rock-over switch for the galvanometer, and the tapping-key was depressed while the average known speed of revolution was maintained. Any advance of a stroboscopically viewed line was immediately rectified by increasing the friction and enforcing its return. Departure from balance was then manifested by the growing drift of the fluxmeter pointer in one or the other direction, and was rectified by adjustment of R until, after a considerable run, the pointer maintained its zero value. This method was proved to be satisfactory by deliberately allowing lines to escape and return in a time small compared with the period of the fluxmeter. Moreover the sensibility was good, the balance of R being sharp on the slide wire, while even small departures from the correct current in the earth coils used for neutralizing the earth's vertical component could readily be detected by this fluxmeter method.

§ 4. EXPERIMENTAL TESTS

Attainment of the linear law $M = K\theta$. The distance pieces between the twin field coils A and B were adjusted by trial, a few readings of the type given in table 1 sufficing until close approximation to the linear law was attained. Symmetry of setting was acquired by throwing A and B into opposition and moving them together until the mutual inductance between them and the rotating coil was as closely

* The recent control-less jewel-pivoted type is less satisfactory owing to solid friction.

as possible zero in all positions. The angle coil likewise was set symmetrically by opposition tests, and its positions of zero mutual inductance with the fixed coils in conjunction were made to agree very closely with the corresponding conjugate positions of the rotating coil. Readings of the mutual inductance M between the rotating coil and the field coils, and the readings of the corresponding mutual inductance M_θ between the angle coil and the field coils, were then taken over all four quadrants at closely corresponding positions. The maximum angle-coil inductance was observed also and the angle θ in seconds of arc was then evaluated from the sine law for all positions. The corresponding values of M and θ in all four quadrants were added together and the ratio M/θ was obtained from $\Sigma M/\Sigma\theta$. The values obtained for the final setting of the field coils for spins with a commutator of small angle are recorded in table 1.

Table 1. Values of M/θ for various angles θ. $x/a = 0{\cdot}247$, $M_{\max.} = 7085{\cdot}17$

ΣM nominal (μH.)	$\Sigma\theta$ (seconds of arc)	θ, approximate (degrees)	M/θ nominal (μH./degree)
2560·6	74,726	5·2	123·35₉
3279·2	95,698	6·6	123·35₈
4033·9	117,726	8·2	123·35₅
4833·2	141,050	9·8	123·35₇
5649·5	164,875	11·45	123·35₅
6448·8	188,209	13·1	123·35₁
7320·8	213,675	14·8	123·34₁
8148·6	237,844	16·5	123·33₇
8947·9	261,176	18·1	123·33₆
9781·3	285,502	19·8	123·33₆
10659·4	311,146	21·6	123·331
12450·2	363,614	25·25	123·265
15043·9	439,795	30·5	123·144

A plot of M/θ against θ reveals peculiarities due to imperfect symmetry, but the approach to linearity is remarkable and the value of K for contacts of approximately known angle can be readily determined with precision.

With the coils in this position the semi-angle of the contacts used in the spin experiments was 11°·6, and accordingly the value of the constant K was taken as 123·355 nominal microhenries per degree. The correcting factor to convert to true microhenries was found with the aid of a standard inductance to be 1·00035, and hence for this arrangement of coils and contacts the experimental value of K is 123·398 μH. per degree of arc.

Spin experiments with small angle of contact. The small commutator was readily adjusted systematically with the aid of the inductometer until the mutual inductance between the rotating coil and the fixed coils had approximately the same value, at the four boundaries of make and break between sectors and brushes, which are crossed during each revolution. The value in question was some 1425 μH., corresponding to a semi-angle of 11°·6. The current in the earth coils was then adjusted, when the rotating coil was spinning with the main current broken, until no deflection was obtained on the galvanometer when closed through R and the commutator of the revolving coil. The speed of revolution was then maintained at a constant value,

determined by the tuning-fork and the stroboscopic disc in use, while R was adjusted until accurate balance was obtained on the galvanometer or fluxmeter when a main current of the order of an ampere traversed the circuit in either direction. The balancing value of R was then immediately checked on the thermoelectric potentiometer by comparing it with one of the standard resistances—usually the 0·50006-Ω. resistance. The value of R in c.g.s. units is given by the expression

$$R = 36 \times 123{\cdot}398 \times 10^4 \times n,$$

where n, the number of revolutions per second, is obtained by dividing the frequency of the fork by the number of lines on the stroboscopic disc. Table 2 gives the results obtained.

Table 2. Absolute measurement of resistance. $K = 123{\cdot}398$ μH./degree

Frequency of fork (c./sec.)	Sectors on disc	n (rev./sec.)	$R \times 10^{-9}$ (c.g.s. units)	Resistance by comparison (Ω.)
256	30	8·533	0·37908	0·37889
320	30	10·667	0·47385	0·47365
384	30	12·800	0·56862	0·56842
384	36	10·667	0·47385	0·47386
512	36	14·222	0·63180	0·63193
384	40	9·600	0·42646	0·42647
512	40	12·800	0·56862	0·56879

The use of the synchronized television motor was limited of necessity to the first three experiments. Fluxmeter balances were taken in the other tests.

Adjustment of twin coils for use with large contacts. The twin coils A and B were now brought closer together and set symmetrically with respect to both the rotating coil and the angle coil by opposition inductometer tests as previously described. The variation of M and θ was then explored over all four quadrants by taking readings of M and M_θ in closely corresponding positions.

On adding the corresponding values of M and θ in the four quadrants we obtain from $\Sigma M/\Sigma\theta$ mean values of M/θ at various angles of displacement. These results are given in table 3 and are well represented by the relationship

$$M/\theta = 124{\cdot}932 + 3{\cdot}665 \times 10^{-4}\,\theta^2 - 3{\cdot}05 \times 10^{-7}\,\theta^4$$

as will be seen from the fifth column, which gives the values calculated from this expression. The differences between the observed and calculated values of M/θ are given in the last column.

This setting of the coils with a maximum value for M/θ of 125·042 nominal μH. per degree at 24°·50 appears ideal for use with a commutator having sector contacts of semi-angle between 23° and 26°. Accordingly the commutator M, figure 2, with smaller sector of angle tapering from 44° to 51° was adjusted on the shaft.

Determination of the constant K' for large angle of contact. The larger commutator was adjusted systematically and firmly fixed in position. The constant K' of equa-

Table 3. Values of M/θ at various angles θ. $x/a=0\cdot212$, $M_{max.}=7249\cdot37$

M (nominal (μH.)	(θ seconds of arc)	θ (degrees)	M/θ observed (nominal μH./degree)	M/θ calculated (nominal μH./degree)	Difference
3261·2	93,961	6·53	124·949	124·947	+0·002
4842·7	139,511	9·69	124·963	124·964	−0·001
7247·3	208,722	14·49	125·000	124·996	+0·004
9690·8	279,048	19·38	125·021	125·027	−0·006
10957·8	315,489	21·91	125·038	125·038	Zero
12255·2	352,824	24·50	125·045	125·042	+0·003
13625·1	392,298	27·24	125·033	125·036	−0·003
14927·4	429,860	29·85	125·014	125·016	−0·002
16834·9	485,007	33·68	124·958	124·955	+0·003

tion (5), equivalent to $(M'/\theta'+M''/\theta'')/2$, was found for it directly *in situ* by taking readings of M and θ at the sector edges over all four quadrants with the brushes (*a*) just on and (*b*) just off. The data are recorded in table 4.

Table 4. Direct determination of K' for contact of 47°

$x/a=0\cdot212$, $M_{max.}=7249\cdot37$

Quadrant	Contact	M (nominal μH.)	M_s (nominal μH.)	θ (seconds of arc)
1	On	2945·7	2890·0	84,579
1	Off	2968·1	2910·7	85,222
2	On	2924·6	2885·9	84,451
2	Off	2927·2	2888·1	84,520
3	On	2880·8	2827·8	82,653
3	Off	2888·3	2835·0	82,876
4	On	2937·2	2898·4	84,840
4	Off	2953·0	2912·5	85,278

The mean value of K' for the on and off positions, which owing to the closeness of the ratios involved is equivalent to $\Sigma M/\Sigma\theta$ for all readings, is $125\cdot040_5$ nominal μH./degree or, when corrected by the factor 1·00035 to convert to true micro-henries, 125·084 μH./degree for a semi-angle of contact of 23°·4.

Spin experiments with an angle of contact of 47°. A series of spin experiments was performed in the manner already described, and the balancing value of R in c.g.s. units was given by

$$R=36\times125\cdot084\times10^4\times n.$$

Table 5 shows the values of R so obtained and the corresponding potentiometer checks. The synchronized motor was used in the first three experiments.

Table 5. Absolute measurement of resistance. $K' = 125{\cdot}084\ \mu$H./degree

Frequency of fork (c./sec.)	Sectors on disc	n (rev./sec.)	$R \times 10^{-9}$ (c.g.s. units)	Resistance by comparison (Ω.)
256	30	8·533	0·38426	0·38398
320	30	10·667	0·48032	0·48018
384	30	12·800	0·57639	0·57614
256	24	10·667	0·48032	0·48008
320	24	13·333	0·60040	0·59982
256	36	7·111	0·32022	0·32047
256	36	14·222	0·64043	0·64036
512	36	14·222	0·64043	0·64026
320	36	8·889	0·40027	0·40002
384	36	10·667	0·48032	0·48033
512	40	12·800	0·57639	0·57627

§ 5. CONCLUDING REMARKS

As this method of measuring resistance absolutely is based on the constancy of the ratio M/θ in the neighbourhood of an angle θ which determines the approximate semi-angle of the contacts employed, the greater part of our time has been given to an investigation of the relationship between M and θ and to devising means of measuring the ratio accurately. Once the laws of inductance made use of have been established, a few readings suffice to determine the constant K for any commutator in position and to set its value between close limits. We favour the use of contacts of some 50° of arc owing to the great accuracy with which K can then be measured.

In view of the high sensitivity and the rapidity and ease with which balance can be obtained, we propose to undertake further work with coils of rather greater diameter (allowing more space between the fixed pair and the rotator) so wound that resistances of the order of an ohm may be conveniently measured. Further investigation is also being undertaken of the limits of accuracy of the sine law for the angle solenoid. It is proposed for the purpose of measuring θ to use two additional twin coils always in the Helmholtz position and possessing a maximum mutual inductance of some 10 mH. with the angle coil. The essential length and time measurements will then be dependent upon the value of a convenient standard inductance and upon a standard tuning-fork frequency.

§ 6. ACKNOWLEDGMENTS

We express our gratitude to Prof. P. M. S. Blackett, M.A., F.R.S. for providing us with facilities for carrying out this research and for the encouragement he has given us throughout the investigation. We are indebted to Mr S. Baker for the design and construction of the fork-controlled motor-synchronizing unit and to Mr H. G. Bell for valuable help in the construction of the apparatus.

DISCUSSION

Dr D. Owen. The method described is not really related to that of the B.A. revolving coil, as appears to be implied in the introduction to the paper. Its affinity is rather with the Lorentz method, as is clearly indicated by the formula $R=Mn$ which applies to both. The idea of using a momentary contact at the instant of maximum induced e.m.f. of the moving conductor was proposed by Lippmann, who used a coil rotating in the uniform field within a solenoid carrying the same current as that passing through the resistance to be measured. The advantage of the present method lies in the application of the investigation previously made by one of the authors of a type of variable mutual inductance in which, over a wide range of angular movement, the mutual inductance is very closely a linear function of the angle. This at once puts the determination of resistance on an altogether higher plane of accuracy. Compared with the Lorentz revolving-disc method, it is now possible to use a multilayered coil, and consequently the scale of size of the whole apparatus, or the speed of rotation of the coil, may be greatly reduced. These advantages may well engage the careful consideration of those concerned with future work on the determination of the ohm at the various national laboratories.

Authors' reply. The method here described resembles the B.A., the Carey Foster and the Lippmann methods in that an alternating e.m.f. is generated in the revolving coil. It differs from them in that the e.m.f. is not sinusoidal and is very uniform over the contacts particularly at the contact edges. In Lorentz's method the magnetic lines of force are cut throughout a revolution at a constant rate and the e.m.f. is unvarying.

We thank Dr Owen for drawing our attention to Lippmann's method, the formula for which may be written $R=Mn$ if the contact is momentary. Our remarks in the paper on the correction for arc of contact in the Carey Foster method are equally applicable to the Lippmann method.

535.62

A NEW PRECISION COLORIMETER

By JOHN J. MANLEY, M.A., D.Sc. Oxon.,
Fellow of Magdalen College, Oxford

Received June 10, 1934. *Read in title November* 2, 1934.

ABSTRACT. In this paper is given a description of a colorimeter (or tintometer) in which use is made of a simple optical system for juxtaposing the two beams of light. Also, novel methods are introduced for equalizing the tints of the two columns of liquid and for determining the difference between their lengths.

§ 1. INTRODUCTION

Of late I have had occasion to carry out many determinations of minute quantities of iron, and for that purpose the precision colorimeter which is the subject of this paper was designed and constructed. The chief features to which attention may be drawn are (i) the simplicity and consequent inexpensiveness of the optical system; (ii) the plan used for varying the length of the column of the standard solution; and (iii) the differential method employed for ascertaining the relative values of the two solutions contained in the tubes *A* and *B*.

§ 2. CONSTRUCTION OF APPARATUS

Figures 1, 2 and 3 are largely self-explanatory. It may be seen that the several glass and other parts of the instrument are, as usual, mounted upon a vertical board *C*. The tubes *A* and *B*, graduated in centimetres, are supported by a shelf *D* having suitable apertures for the transmission of light reflected from the mirror *M**. Variations in the inclination of the mirror are producible by means of the screw *S*. On emerging from the tubes *A*, *B*, the two beams of light enter and pass through the V-shaped block of glass *G*, *G*, and are thus brought close together. Each limb of the vee is 18 mm. thick and the faces are polished and parallel. The two plates, united with Canada balsam, meet and form as shown in the figure an angle of 78°. After traversing the prismatic block the beam of light is, by means of a diaphragm *F*, *F* having an opening 4 mm. long and 3 mm. wide, so limited that the central portions only pass upwards into the eye-piece *E*. The level of the standard solution in *B* can be varied and adjusted by a glass plunger *P* contained within a cistern *T*. The plunger, of uniform diameter, is suspended from a pin *N* by a platinum ribbon *O*

* The tubes *A* and *B* are not enclosed. Theory and experiment alike show that the use of a cover to exclude light other than that reflected by the mirror is wholly unnecessary. A decided advantage, however, accrues from the screening of the prismatic block for which an aluminium cover is used.

0·1 mm. wide and 0·08 mm. thick; and it is so loaded with mercury that even when it is immersed to the fullest extent the ribbon still remains taut. For keeping the plunger central three solid glass pimples *a*, *b*, *c* were formed and then attached by fusion as shown. In the absence of these the plunger makes contact with the walls of the cistern, thus causing inconvenience and inaccuracy. Two stops *d*, *e* impose a limit to the degree of rotation of the disc *X* carrying the index *I*. A ring *L* having upon its outer rim a millimetre scale as shown can be rotated independently of *X*

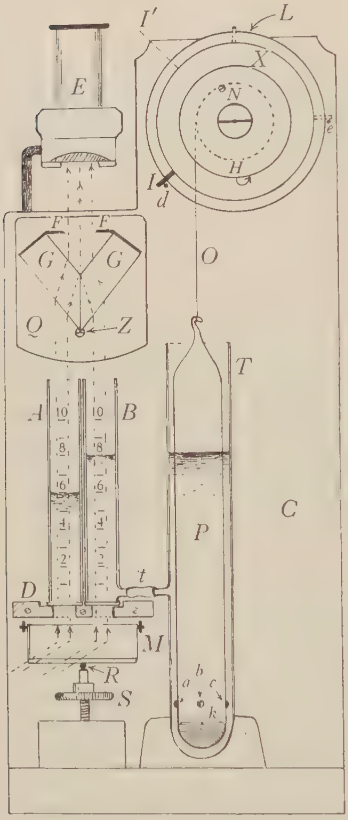


Figure 1.

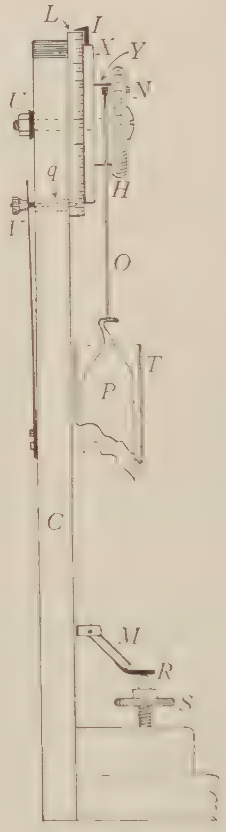


Figure 2.

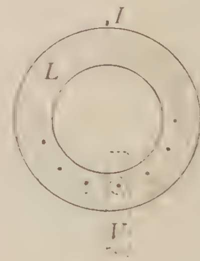


Figure 3.

and then secured in any one of a number of positions by a pin *q* fixed to a steel spring *V*. For convenience, the diameter of the cylinder *Y* and that of the graduated ring *L* are so proportioned that when *X* is rotated and *P* raised or lowered until the level of the solution in *B* has been changed by $\mp$ 10 mm., the reading as shown by *I* is altered by $\pm$ 20 scale divisions. Under these circumstances each scale division (1 mm.) corresponds to a variation in the level of the liquid in *B* equal to $\mp$ 0·5 mm. Smaller quantities are a matter of estimation. The plunger is retained by friction in any position to which it may be brought, and the friction, which is but small, is produced by an appropriate adjustment of the nut *U*.

In figure 3 is shown the posterior surface of the ring *L* with its series of equally spaced cavities* for the reception of the pin *q*. The distance between any two cavities is such that it is equivalent to a difference of 0·2 unit (cm.) of the scale upon either glass tube, and to 4 scale divisions of the ring *L*.

The plate *Q* is pivoted upon the pin *Z*, and thus any required adjustment of the prismatic block about a horizontal axis can be readily effected; the plate can then be clamped by means of a nut (not shown) at the back of the support.

For attaching the parallel-plane plates to the tubes *A* and *B*, use was made of a cement sold under the name of "Durofix." This cement, which is not affected by water, has proved highly satisfactory.

§ 3. EXPERIMENTAL PROCEDURE

The colorimeter is used in the following way. First, the coloured solution under examination is introduced into *A* and the level of the liquid adjusted so that it coincides with an appropriate division, say 3, 6 or 8. The most convenient length for the column obviously depends upon the intensity or depth of the colour. Let us suppose the length to be 5 cm. Secondly, the plunger *P* is moved by turning *H* until the index *I* is in the position indicated by *I'*. Thirdly, *V* is now pulled back so as to withdraw the attached pin *q* from its cavity: then the graduated ring *L* is rotated until the zero of its mm. scale is in close proximity to the index at *I'*. On releasing *V* and slowly turning *L* to the right or left *q* springs into an adjacent cavity and thus for the time being maintains the ring in a fixed position. The index is then set at zero. Fourthly, the standard solution is poured into the cistern *T* until the level of the liquid in *B* coincides with the graduation 4 (not with 5 as in the case of *A*). By adopting this plan, the level of the liquid in *B* can by means of the plunger be varied within the limits $5+2$ and $5-2$ cm. Lastly, by means of the milled head *H* the plunger is alternately raised and lowered until finally the two patches of colour as seen through the eye-piece are precisely matched. The index reading is then noted, and as this at once gives the difference in the levels of the two columns the required value may then be calculated.

It will be observed that differences measured in the way just described assume values which are positive when the plunger is lowered and negative when it is raised. Accordingly all graduations of the scale on the left of zero as seen from the front are marked +, whilst those on the right are marked −.

It may be desirable to point out that if some level other than 5 be more convenient for the measurement in hand the procedure remains unchanged. For example, let the length of the column in *A* be increased to 7. This column is 2 cm. longer than the former, and as each 1 cm. is equivalent to 20 divisions of the graduated ring, the ring must be rotated through 2×20 to the right and then, as in the first case, secured by the pin *q*. Next, the index *I* having been brought to the zero, an additional quantity of the standard solution is introduced into the cistern *T* and the level of the liquid in *B* is adjusted so as to coincide with the graduation marked

* Actually the cavities are five times more numerous than those represented in the figure.

6. That is with 7 - 1 instead of with 5 - 1 as in our first illustration. Finally the two colours are matched as already described and their relative values calculated.

When the instrument is to be used for liquids which are affected by rubber the side-tubes at *t* may be united by fusion.

§ 4. ACCURACY ATTAINABLE

To those experienced in the practice of colorimetry, it will be quite obvious that the degree of accuracy attainable with the instrument just described must be, and is, of the same order as that ensured by the use of other precision colorimeters. And here we remark that the chief difficulty associated with the use of colorimeters arises not so much from the form of any given instrument as from a lack, on the part of the observer, of a necessary sensitivity for apprehending differences in colour-density. Some observers are able to match particular colours with ease but experience considerable difficulty when dealing with others. In this respect my own observations are uncertain and irregular in the case of yellow, pink or red solutions, but are reasonably uniform for those which are blue. And so for the research which, as we have already stated, involved the determination of minute quantities of iron, choice was made of Prussian blue rather than of the red compound iron-potassium thiocyanate. Curiously enough a friend engaged in a similar research prefers the latter to the former compound.

For ascertaining the mean error attending the use of the new instrument, the following experiments were made. First an acid solution of iron in the ferric state was prepared, the concentration of which was 56 mg. of iron per 100 cm.³ Of this solution 1 cm.³ was transferred to a 100 cm.³ measuring-flask and freely diluted. Then the required potassium ferrocyanide was added, and the whole was well mixed and finally made up to the standard volume. Newly prepared, such a dilute solution remains quite unclouded for some days.

Next, both tubes of the apparatus were charged with standard solution and the necessary settings and adjustments carried out in the way already described. For the first experiment the column of liquid in *A* was 40 mm. long. Finally, column *A* was repeatedly measured in terms of column *B* with the results set forth in table 1.

Table 1

Length of column (mm.)			
Observed		True	Differences (mm.)
Plunger raised	39·5	40·0	−0·5
	40·5	40·0	+0·5
Plunger lowered	39·5	40·0	−0·5
	41·0	40·0	+1·0
Mean = $40·1_3$			Mean = $+0·1_3$

The mean difference $+0·1_3$ is equivalent to +0·065 mm. (*vide supra*). Hence the measured weight of iron differs from that known to be present by $0·56 - \frac{0·56 \times 40}{40·065}$ or $0·56 - 0·559_9$ mg.

In continuation of the series of tests, *A* and *B* were again charged with the solution containing 0·56 mg. of iron per 100 cm.^{3} The procedure detailed above was again followed, and other 3 columns of the standard solution were measured with the results set forth in table 2.

Table 2

Length of column in tube *A* (mm.)	Mean measured length of column in tube *B* (mm.)	Iron content		
		True (mg.)	Observed (mg.)	Differences (per cent)
60	59·62	0·56	$0{\cdot}5564_5$	−0·64
80	79·88	0·56	$0{\cdot}5591_7$	−0·14
100	100·01	0·56	$0{\cdot}5600_6$	+0·04
			Means = 0·5586	−0·25

Finally, with the object of rendering these tests more complete, several solutions having concentrations less than that of the standard liquid were prepared and measured with the results shown in table 3.

Table 3

Solution no.	Length of column in *A* (mm.)	Iron content		
		True (mg.)	Found (mg.)	Difference (per cent)
1	90	0·187	0·1865	−0·27
2	90	0·250	0·2485	−0·60
3	60	0·280	0·2770	−1·08
4	60	0·373	0·3720	−0·27
				Mean = −0·56

From tables 2 and 3 it may be seen that the final mean error is of a negative character and that therefore, in general, the weight of the iron was underestimated. But given another observer, having a personal equation differing from my own, the corresponding mean error might well be dissimilar not only in magnitude but also in sign.

§ 5. EFFECTS OF EYE-FATIGUE

In conclusion, attention may be drawn to the way in which the effects of eye-fatigue resulting from an indispensable critical observation of the colour-patches were met and overcome. Experiment showed that for a series of readings the greatest uniformity in apparent values was obtained when the two columns were balanced quickly, or by first approximating to the true value and then resting the eye before the final adjustment was made. Another and most helpful plan was that in which the main adjustment was carried out with the aid of one eye and the final and more precise setting accomplished with the aid of the other and unfatigued eye. The apparent increase in the brightness of the field, as judged by the second eye, is usually very considerable, and so any residual difference in the depth of colour of the two columns is the more readily detected and balanced. I am not aware that these important matters of detail have hitherto received the notice they seem to merit.

537·565 : 546.293

THE ENERGY OF AGITATION OF POSITIVE IONS IN ARGON

By F. LLEWELLYN JONES, M.A., D.Phil.,
University College of Swansea

Communicated by Prof. E. J. Evans, July 2, 1934. Read in title, November 16, 1934.

ABSTRACT. The energy of agitation of positive argon ions was investigated for a range of electric forces Z from 5 to 50 V./cm. and for a range of pressures p from 3 to 0·24 mm. by means of the Townsend method of measuring the lateral diffusion of a beam of ions moving in a uniform electric field.

In pure specimens of gas the energy of agitation of the ions was the same as that of the atoms of the gas, and independent of the pressure, when the ratio Z/p was less than about 20. For higher values of Z/p the mean energy was increased by the field; the values obtained are compared with those calculated on the assumption that the ions behave as perfectly elastic spheres which attract the atoms of the gas with a force varying as the inverse fifth power of the distance.

The motion of the ions was found to be very sensitive to the presence of impurities, and the general effect of the impurities at pressures greater than 0·4 mm. and when Z/p was less than 80 was to reduce the mean energy of agitation of the ions to the value possessed by the atoms of the gas. This is probably due to the formation of ionic clusters.

§ 1. INTRODUCTION

p
Z

When a stream of positive ions moves through a gas at a pressure of p mm. in a uniform electric field Z their mean energy of agitation is not very different from that of the molecules of the gas when the ratio Z/p is less than about 10*. When Z/p is increased, there are various processes which determine whether the mean energy of the ions also increases. The influence of these processes has been fully discussed by Tyndall and Powell† in their investigation of the mobilities of positive ions in pure gases. If the gas contains impurities, molecules of the impurity may collect around the monatomic ion to form a cluster of many times atomic magnitude. This would especially be the case at high gas pressures and if the impurity molecules have a permanent dipole. Owing to their large mass the mean energy of agitation of these clusters would then be the same as that of the atoms of the gas, even for higher values of the ratio Z/p of the order of 100. However, as the pressure is reduced the influence of impurities decreases rapidly. The chance of a cluster being formed will decrease, and, as the electric intensity is increased the clusters will be broken up by collisions with atoms, and the ion will become monatomic. Again, the process of electron-transfer from a neutral atom to the positive ion with which it collides is considered by some physicists to modify

* Z expressed in V./cm.

† A. M. Tyndall and C. F. Powell, *Proc. roy. Soc.* A, **129**, 162 (1930); A, **134**, 125 (1931).

the energy of agitation of the ions, owing to the fact that the charge may not remain associated with the same ion as it moves through the gas*. This process is the more important in the case when the ions are those of the gas through which they move. It is suggested† that electron-transfer introduces short-range interaction forces between the ions and atoms of the gas, so that the total force between a positive ion and gas atoms must contain these exchange forces in addition to those due to the polarization of the neutral atoms by the charge on the ions. The mobility of the ions, and hence their energy of agitation, depends on these forces. Hassé and Cook‡ have shown, however, that if the collisions between the ions and atoms resemble those between elastic spheres the transfer of charge makes no difference to the mobility of the ions, but in general, when the ion exerts forces on the atoms, the mobility is reduced by the process.

§ 2. DESCRIPTION OF APPARATUS

In order to investigate the variation of the mean energy of agitation of positive ions with the ratio Z/p the following experiments were performed in argon. The best method for the determination of the energy of agitation of ions in gases in uniform electric fields is that due to Townsend§, which consists in measuring the lateral diffusion of a narrow beam of ions moving in a uniform electric field between two parallel plates.

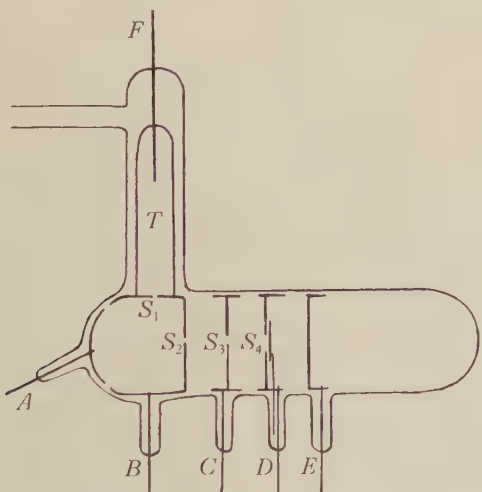


Figure 1.

The apparatus employed is shown diagrammatically in figure 1. The design of the electrodes C, D and E was similar to that used in an apparatus for measuring

* F. M. Penning and C. Veenemans, *Z. Physik.* **62**, 746 (1930).
† J. E. Lennard-Jones, *Proc. phys. Soc.* **43**, 461 (1931).
‡ H. R. Hassé and W. R. Cook, *Phil. Mag.* **12**, 554 (1931).
§ J. S. Townsend, *Proc. roy. Soc.* A, **81**, 464 (1908) and *Motions of Electrons in Gases*, Clarendon Press, pp. 7–9.

the ionization produced by positive ions described previously*. They were thin discs 3·4 cm. in diameter, fitted with cylindrical rings, and mounted 1·8 cm. apart on three pyrex rods. The slit S_3 in the electrode C was 1·25 mm. wide and 5 mm. long, and the slit S_4 in D was 2·5 mm. wide and 6·5 mm. long. A shutter forming part of the electrode D could be placed over the slit or withdrawn to the side to leave it open, and was operated magnetically.

The positive ions which pass through these slits were generated in a side tube by means of a direct-current discharge between the molybdenum wire anode F and the electrode B. A fraction of the ions which passed through the slit S_1 were repelled by the electrode A and passed through the slit S_2 and finally entered the space between C and D. Slits S_2, S_3 and S_4 were in the same straight line. Ions were prevented from reaching the electrode C except through the slit S_2 by confining the discharge within a glass tube T, 6 cm. long and 1·5 cm. in diameter, which fitted tightly into a collar fixed to the cathode. Also, no ultraviolet light from the discharge in T could pass through the slits S_3 and S_4. The electrodes were made of copper 0·5 mm. thick and were insulated by pyrex bars. The glass envelope was also of pyrex and the connections to the electrodes were brought out through molybdenum seals.

The electrode C was maintained at earth potential, and the potential difference between A and B was about 10 volts. The electric force Z between the electrodes E and D was the same as that between D and C, which was approximately the same as that between the electrodes C and B. Under these conditions the positive ions attained the mean energy determined by the ratio Z/p before entering the slit S_3. The potentials of the electrodes were obtained from a battery of small accumulators.

The source of high potential for the discharge tube was a 600-volt direct-current generator. The current could be adjusted to any desired value by means of a diode valve interposed in the positive lead from the generator. Thus the intensity of the current of positive ions was controlled by the filament rheostat of the diode, and their mean energy of agitation was determined by the potential-difference between the electrodes B and C. The simplest method of cutting off the stream of positive ions was found to be by means of a single-pole switch placed in the anode lead to the discharge tube in series with a high resistance of the order of 1 MΩ. In this way the difference between the sparking and the maintenance potentials of the discharge was set up across the resistance, and the discharge current always attained the same value on closing of the switch. Thus the apparatus was readily adjusted to give any required positive ion current to the collectors.

§ 3. PURIFICATION OF GAS

The experiments were made with argon which was obtained commercially pure from the General Electric Company, Ltd. Further purification was effected by circulating the gas over red-hot calcium, in the manner devised by MacCallum and Klatzow†. The calcium was contained in a silica tube which could be raised

* J. S. Townsend and F. Llewellyn Jones, *Phil. Mag.* **15**, 282 (1933).
† S. P. MacCallum and L. Klatzow, *Phil. Mag.* **15**, 829 (1933).

to red heat electrically. This tube was connected to the apparatus by means of a ground-glass joint, vacuum grease being used on only the outer half of the seal. The diffusion of mercury vapour from the Mcleod gauge and vapour pump into the pure gas was prevented by means of a liquid air-trap in the usual manner.

The pyrex apparatus was heated for several hours to expel gases from the electrodes and from the surface of the glass, and a discharge was also maintained in the side tube. The argon was kept in contact with the red-hot calcium for some time before entering the apparatus. The gas was then withdrawn by absorbing it in a charcoal trap cooled by liquid air, again passing over the heated calcium. This operation was repeated many times and the argon was stored over the calcium while it cooled to room-temperature before being finally admitted to the apparatus for the measurement of the currents. The circulation of the gas over the hot calcium was also repeated during the course of the experiments.

§ 4. EXPERIMENTAL PROCEDURE

The discharge current between the electrodes A and B had a value between 60 and 120 μA., depending on the pressure of the gas, while the currents of positive ions passing through the slit S_3 were of the order of 10^{-11} A. Larger currents are undesirable in order that the space charge may not distort the electric field between the electrodes. Under these conditions the currents were measured by observing the charges received by the electrodes D and E when a discharge was maintained for 10 seconds in the side tube. The charges were measured by an electrostatic induction balance consisting of a quadrant electrometer, screened air condensers, and a potentiometer, in order that the potential of the insulated electrode may not change while the electrode was receiving a current of positive ions.

At some pressures it was found that a visible stream of positive ions passed through S_1 into the space between A and B. Some of the ultra-violet light emitted passed through the slit in the electrode C and caused the emission of electrons from D and E. In order to estimate the currents due to the positive ions it is necessary to correct the observations by deducting from the charges received by the discs D and E the small charges received when the electrodes B and C were at the same potential, for then the positive-ion current passing through the slit S_3 was negligible. These corrections were of the order of 5 per cent of the positive ion current at the higher pressures but rose to 20 per cent at the lowest pressure with the higher forces.

§ 5. RESULTS OBTAINED

As the ions move from C to E under the electric force Z the beam diverges so that some fall on the electrode D, and the rest, when the slit S_4 is open, pass through to be collected on E. When the shutter closes the slit, the whole beam of ions is collected on electrode D. The experimental part of the investigation consisted in measuring the currents i_1 to the electrode E, the current i_2 to the electrode D when the slit was open, and the current i_3 to D when it was shut. If all the ions which

pass through the slit are collected on E, then $i_3 = i_1 + i_2$; and this was found to be the case in all the measurements. The ratio i_1/i_3 was determined at various pressures of gas from 3 mm. to 0·24 mm. for a range of electric forces from 10 to 50 V./cm.

The results of the experiments with the purest specimens of argon are shown
R by the curves in figure 2, in which the ordinates represent R, the ratio i_1/i_3, while the abscissae represent the electric intensity.

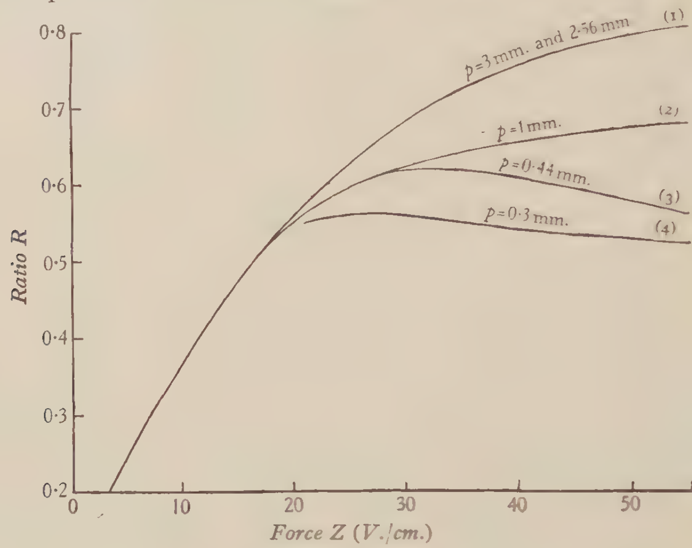


Figure 2.

Curve (1), obtained at a pressure of 3 mm., may be taken to correspond to the case when the energy of the ions is the same as that of the atoms of the gas, the ratio Z/p being less than 17. When the pressure was reduced to 1 mm. curve (2) was obtained, and for values of Z/p less than 20 it is identical with curve (1). This shows that values of Z/p less than about 20 have little effect in increasing the energy of agitation of the positive ions. However, for higher values of Z the divergence of the ions increased, indicating an increase in the mean energy. Curve (4), obtained at a pressure of 0·3 mm., shows that the energy of agitation was increased still further, and the ratio R at about 0·5 was independent of higher values of the electric force.

However, during the course of the experiments it was found that the presence of minute quantities of impurities had a large effect on the energy of the positive ions. Thus at a pressure of 1 mm. a specimen of gas was allowed to remain in the apparatus and was not circulated over the calcium during the measurement of the currents. The curve thus obtained practically coincided with curve (1), showing that the energy of agitation of the ions was the same as that of the atoms of the gas for values of Z/p up to 60. Highly purified argon has been prepared by MacCallum and Klatzow*, who found that the presence of impurities in minute quantities has a large effect on the sparking-potential and the electrical properties of the gas.

* *Loc. cit.*

Experiments were also performed with positive ions in specimens of argon which were known to be impure. Purified argon at a pressure of 1·52 mm. was admitted to the apparatus and frequent discharges were produced in the side tube, currents of 200 μA. being used in order to cause the emission of impurities. The gas was not circulated over the heated calcium for a period of many weeks. Measurements of the ratio R were made for a range of pressures between 1·52 mm. and 0·24 mm. The results are given in the curves of figure 3.

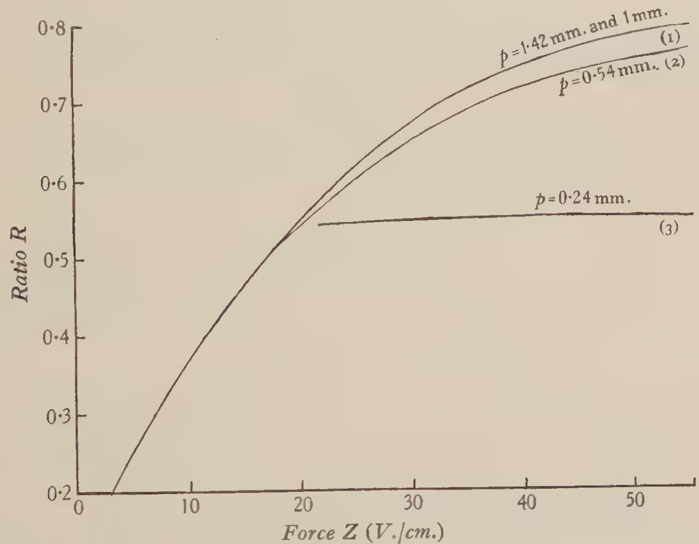


Figure 3.

For pressures of 1·5, 1·0 and 0·54 mm. the curves are all very close together, all roughly corresponding to the case when the energy of the ions was equal to that of the gas atoms. Thus the presence of the impurities appears to prevent the energy of agitation of the ions from becoming greater than that of the atoms of the gas for pressures greater than about 0·5 mm. with electric forces less than 60 V./cm. However, when the pressure was reduced to 0·24 mm. the energy of the ions was increased by the electric field just as in the case of the purer specimens of gas. A reasonable explanation of this effect is the increase in the mass of the ions due to the clustering of molecules of the impurity around them; and the experiments show that these clusters may exist even when the ratio Z/p is about 80. It would thus appear that impurities have much less effect at the low pressures of the order of 0·25 mm. in reducing the energy of agitation of the ions. This is doubtless due to the smaller chance of the formation of clusters, and the strong electric field breaking them up.

By comparing the curves for a pressure of 1 mm. obtained with the pure and impure specimens of argon it is seen that the effects produced by the impurities appear not to depend very greatly on the quantity of the impurity present. Thus the curve for argon containing only traces of impurities—approximately curve (1),

figure 2—is practically the same as the curve obtained with the very impure specimens, curve (2), figure 3. The effect due to the first trace of impurity is very great. The high sensitivity to impurities of the motion of the positive ions in argon makes it probable that the values of the ratio shown by curves (2), (3), (4) in figure 2 are too high, that is that the electric field increases the mean energy of agitation of the ions more than is indicated in these experiments.

§ 6. INTERPRETATION OF RESULTS

Townsend's* well-known analysis of the motion of ions in a uniform electric field Z V./cm. shows that the ratio R is given by the equation

$$R = f\,(W/D) \qquad \text{......(1),}$$

W, D — where W is the velocity of the ions in the direction of the electric force and D their coefficient of diffusion, supposed isotropic. But

$$\frac{W}{D} = \frac{3Z}{2E_1} = \frac{3}{2E_2}\cdot\frac{Z}{k} \qquad \text{......(2),}$$

E_1; k, E_2 — where E_1 is the mean energy of agitation of the ions expressed in volts which is equal to kE_2, E_2 being the energy of agitation of the gas atoms in volts. Thus

$$R = f\,(Z/k) \qquad \text{......(3).}$$

When the energy of agitation of the ions is not very different from that of the atoms of the gas, Z/k is equal to Z, and the theory shows that R is a function of Z and is independent of the pressure of the gas. Thus for pressures greater than about 1 mm. and for values of Z/p less than about 20 the curves are identical.

When the electric intensity is very high and the ionic mass is equal to that of an atom of the gas, the drift velocity W of the ions may become comparable with (u) their velocity of agitation u. Consequently the equations of diffusion would be invalidated and equation (3) would require modification. In this case the velocity of the ions is principally in the direction of the electric force, and their energy is due to this motion and not to their motion of agitation. Hence in these experiments the ratio Z/p was limited to values less than 150. For large values of Z the free paths of the ions are bent appreciably in the direction of the field, and the lateral diffusion of the ions is reduced. Thus there is an apparent reduction in the coefficient of lateral diffusion D, and since the energy of agitation of the ions is obtained from this coefficient by means of the equations (1) and (2), it follows that the values of k calculated from equation (3) are too low when Z/p is very large.

An approximately correct modification of the equations may be obtained simply in the case when the collisions between the ions and atoms resemble those between elastic spheres.

In the absence of the electric force let OA, figure 4, represent the free path of (τ) an ion after a collision at O. The mean interval, τ, between successive collisions is ($\lambda, \bar{u}$) equal to $\lambda/\bar{u}$, where λ is the mean free path and $\bar{u}$ the mean velocity of agitation. When a strong field Z is acting, in the absence of a collision during an interval τ (s) the ion traverses the parabolic path OCB of length s.

* *Loc. cit.*

Since the mean free path is practically unaltered, the ion now collides at C, where $OC=\lambda$. The lateral diffusion of the ion is thus reduced from OY to OY'. The fractional reduction f is OY'/OY, which, though less than OC/OB, is practically equal to it, i.e. to λ/s. If $\bar{u}_1$ is the mean velocity with which the ion traverses the parabolic path s, then $s=\bar{u}_1\tau$; hence

$$\lambda/s=\bar{u}/\bar{u}_1.$$

Now
$$\bar{u}_1=[\bar{u}^2\sin^2\theta+(\bar{u}\cos\theta+\tfrac{1}{2}\gamma\tau)^2]^{\frac{1}{2}},$$

where $\gamma=eZ/m$, e being the ionic charge and m the mass of the ion. Thus

$$f=[1+\gamma\tau\cos\theta/\bar{u}+(\tfrac{1}{2}\gamma\tau/\bar{u})^2]^{-\frac{1}{2}}.$$

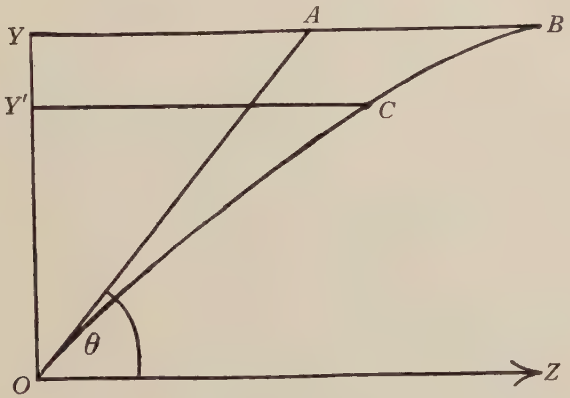


Figure 4.

When the velocity of the ions in the direction of the electric force is large, no great error is made in replacing the mean value of $\cos\theta$ by unity, so that $\bar{f}$ is given by

$$\bar{f}=\frac{1}{1+\frac{1}{2}\gamma\tau/\bar{u}}=1\Big/\left(1+\frac{Z}{p}\cdot\frac{\lambda_1}{4kE_2}\right),$$

where λ_1 is the mean free path of the ions at a pressure of 1 mm. and is computed below* as a function of k, equation (5). The coefficient D_1 of lateral diffusion of the ion is proportional to the mean free path of the ions in a direction perpendicular to the electric force and is thus given by

$$D_1=D/(1+C.Z/p) \qquad \text{......(4)},$$

where C is a function of k, and is $4{\cdot}8\times10^{-3}$ when k is 5. Thus the correction necessary is less than about 30 per cent when Z/p is 60.

It may be seen† that when the coefficients of diffusion along and perpendicular to the electric force are not the same that equation (1) becomes

$$R=f(W/D_1).$$

* See p. 82.
† See for example V. A. Bailey, *Phil. Mag.* 9, 560 (1930).

f
$\bar{u}_1$
γ, e, m
λ_1
D_1
C

Hence the values of k obtained from (3) should be multiplied by $(1+C.Z/p)$ to obtain the energy of the ions. These corrected values of k are given in the table for the different values of the ratio Z/p used in the experiments.

Table 1

k	1	1	1·5	2	2·7	3·8	5
Z/p observed	0	18	27	45	75	100	150
Z/p calculated from equation (10)	0	14	30	43	60	80	100

§ 7. THEORETICAL CONSIDERATIONS

The energy of agitation acquired by the positive ions in a gas under the influence of the electric field depends on the mean free path and on the fractional loss of energy in an encounter with a gas atom. In the steady state, when the mean energy of agitation remains constant as the ions move through the gas, the energy acquired from the electric field is dissipated in collision with the atoms. In order to calculate
E_1 this mean energy E_1 exactly it is necessary to know the nature of the forces between the ions and atoms. On the assumption that the ions exert a repulsive force on the atoms varying as the inverse fifth power of the distance, Pidduck* has shown that k is related to W by the equation

$$k-1=W^2/\Omega^2,$$

Ω where Ω is the velocity of agitation of the gas atoms. However, it is interesting to estimate the energy of the ions on the assumption that their collisions with gas atoms resemble those between elastic spheres, and that in addition they attract the atoms with the force due to the polarization of the gas atoms by the ions. In this case the attractive force between the ion and an atom due to the charge induced on the atom has been given by Langevin† in the form

$$F(r)=\frac{(K-1)\,e^2}{2\pi N r^5}=\frac{B}{r^5},$$

r, K where r is the distance between the centres of the atom and the ion and K the
N dielectric constant of the gas containing N atoms per unit volume. The rate of loss of energy of ions in a gas under these conditions has been investigated by J. J. Thomson‡ who showed that the increase in the energy-loss due to the charge on the ions may be regarded as being equivalent to a reduction in their effective mean
λ, λ' free path. Let λ be the mean free path of the ions and λ' the mean free path of the atoms with the same energy. Then if the attractive force is strong

$$\lambda=\frac{\lambda'}{4{\cdot}4\,(\omega/E_r)^{\frac{1}{2}}} \qquad \text{......(5)},$$

* F. B. Pidduck, *Proc. roy. Soc.* A, **88**, 296 (1913).
† P. Langevin, *Ann. Chim.* (*Phys.*), **8**, 245 (1905).
‡ Sir J. J. Thomson, *Phil. Mag.* **47**, 337 (1924).

where ω is the potential energy of an atom in contact with an ion, and E_r is the relative kinetic energy of the ion. Thus ω, E_r

$$\frac{\omega}{E_r}=\frac{B}{4\sigma_{12}{}^2E_r},$$

where σ_{12} is the sum of the radii of an atom and an ion. σ_{12}

In deducing this relation it is assumed that the force on the atoms of the gas due to the charge on the ions is strong enough to bend the paths of the ions or atoms so that collisions between them are directly along the line of centres, and also that exchanges of momentum occur between the ions and the atoms when they do not collide. In general the rate of loss of energy by the ions is less than that deduced from these assumptions, but the error introduced in adopting equation (5) becomes smaller the more polarizable are the atoms of the gas.

When the positive ions are not in thermal equilibrium with the gas atoms, their mobility μ may be found from their coefficient of diffusion D by means of Langevin's relation given in equation (2). If the velocities of the ions are distributed according to Maxwell's law, D is given by the equation* μ, D

$$D=\frac{4}{(3\pi)^{\frac{3}{2}}N\sigma_{12}{}^2}\left(\frac{E_1}{m_1}+\frac{E_2}{m_2}\right)^{\frac{1}{2}} \qquad \text{......(6)},$$

E_1 and E_2 being the mean energies of the ions and atoms measured in ergs. When the ions are those of the gas through which they move, then $m_1=m_2=m$ the atomic mass and $\sigma_{12}=\sigma$ the atomic diameter. E_1, E_2 σ

The mean free path λ' of an atom with energy E_1 is given by the equation

$$\lambda'=\frac{1}{N\pi\sigma^2(1+E_2/E_1)^{\frac{1}{2}}}=\frac{l\sqrt{2}}{(1+E_2/E_1)^{\frac{1}{2}}} \qquad \text{......(7)},$$

where l is the mean free path of atoms of the gas with mean energy E_2. The mobility μ of the ions is thus given by† l

$$\mu=\frac{0{\cdot}92el_1(1+E_2/E_1)^{\frac{1}{2}}}{(mE_1)^{\frac{1}{2}}} \qquad \text{......(8)},$$

where $$l_1=\frac{l}{4{\cdot}4\,(\omega/E_r)^{\frac{1}{2}}}.$$

The mean fractional loss of energy F in a collision between equal elastic spheres

* Jeans, *Dynamical Theory of Gases*, Chap. x, p. 316.

† It is interesting to note that when the ions are in thermal equilibrium with the gas atoms equation (8) may be expressed in the form given by L. B. Loeb, *Phil. Mag.* 49, 518 (1925), thus

$$\mu=A\left[\frac{2}{\rho\,(K-1)}\right]^{\frac{1}{2}},$$

where ρ is the density of the gas, and A is 0·25. This constant is about one half of the value given in the rigorous derivation of Langevin (*loc. cit.*) which agrees with the experimental determinations for the alkali ions in argon made by A. M. Tyndall and C. F. Powell, *Proc. roy. Soc. Lond.* A, 136, 145 (1932). This difference is due to the adoption of the simple expression (5) for the change in the mean free path due to the polarization of the gas. ρ, A

when the distribution of their energies about their mean energies E_1 and E_2 is Maxwellian is*

$$\tfrac{2}{3}(1 - E_2/E_1).$$

The average number of collisions made per second by an ion of energy $E_1 = \frac{1}{2}mu^2$ may be taken to be $\bar{u}/\lambda$, i.e. $0{\cdot}92u/\lambda$. Hence if E_1 is maintained approximately constant the average rate of loss of energy in collisions is equal to

$$0{\cdot}92FE_1u/\lambda.$$

In statistical equilibrium this loss is balanced by the energy, $e\mu Z^2$, gained per second from the electric field. Thus

$$0{\cdot}92FE_1u/\lambda = e\mu Z^2,$$

and from equations (5), (7) and (8), since $E_r = \frac{1}{2}E_2(1+k)$, then

$$k\left[\frac{1-1/k}{1+k}\right]^{\frac{1}{2}} = \frac{0{\cdot}2elZ}{E_2(\omega/E_2)^{\frac{1}{2}}} \qquad \text{......(9).}$$

Now if Z is measured in volts per centimetre and E_1 and E_2 in volts, then E_2 is $\frac{1}{27}$ V.
L Also $l = L/p$, where L is the mean free path of argon atoms at a pressure of 1 mm. The value of σ deduced† from the coefficient of viscosity is $2{\cdot}96 \times 10^{-8}$ cm. giving $L = 7 \times 10^{-3}$ cm. Again, since $(k-1)$ for argon is $5{\cdot}29 \times 10^{-4}$ it follows that B is $7{\cdot}48 \times 10^{-43}$, so that ω is $2{\cdot}3 \times 10^{-13}$ ergs and E_2 is 6×10^{-14} ergs; thus $(\omega/E_2)^{\frac{1}{2}}$ is 2. Hence equation (9) becomes

$$\left[\frac{k(k-1)}{k+1}\right]^{\frac{1}{2}} = 1{\cdot}88 \times 10^{-2}\frac{Z}{p} \qquad \text{......(10).}$$

The values of the ratio Z/p corresponding to various values of k obtained from this equation are given in the table, where they may be compared with the experimental results.

§ 8. CONCLUSIONS

It has been indicated in § 6 that the experimental values of k may be smaller than the actual values owing to the action of impurities, but for the highest values of Z/p the influence of impurities is not so important. Furthermore, the calculated values of Z/p were obtained on the assumption of the predominating influence of the polarization of the gas atoms, so that they must be regarded as upper limits to these values.

It will be seen then that although a general agreement obtains between the observed and calculated values when Z/p is less than about 80, there is a difference between them which increases when the ratio becomes higher. This difference may be due to the neglect of two considerations: (i) the mean free path of the positive ions at the velocities considered may be shorter than that deduced by equations (5) and (7) from the kinetic theory of gases; and (ii) exchange processes occurring in some of the collisions tend to diminish the mean energy of agitation of the ions. Now the mean free path of positive ions in argon has been found, at certain

* M. Cravath, *Phys. Rev.* **36**, 248 (1930).
† H. R. Hassé and W. R. Cook, *Phil. Mag.* **3**, 977 (1927).

velocities, to be very different from that deduced on the kinetic theory of gases; an effect which is similar to the Townsend-Ramsauer effect for slow electrons. Thus Thomson* found a variation of the mean free path of protons in argon, the minimum occurring at the same velocity as with electrons; while Ramsauer and Beeck† have shown that the cross-sections of argon atoms in collision with the alkali ions increased rapidly when the energy of the ions was reduced to the order of 1 V. Such variations of the mean free path of the ions would cause corresponding variations in the mean energies of the ions.

Again, Hassé and Cook‡ have shown that the process of electron exchange reduces the mobility of positive ions by an amount which depends on the probability of a transfer occurring in any collision with an atom. Since the energy acquired by the ions in moving through the gas is determined by their velocity in the direction of the electric force, exchange processes probably tend to diminish the mean energy of the ions.

It was not possible to determine the relative importance of these considerations in the experiments described in this paper, and the investigations are being extended to measure the coefficients of ionization by positive ions of argon under strong electric forces.

§ 9. ACKNOWLEDGMENT

In conclusion I wish to thank Prof. E. J. Evans for extending the facilities of his laboratory and for his kind interest.

* G. P. Thomson, *Phil. Mag.* **2**, 1076 (1926).
† C. Ramsauer and O. Beeck, *Ann. Phys.*, Lpz., **87**, 1 (1928).
‡ H. R. Hassé and W. R. Cook, *Phil. Mag.* **12**, 554 (1931).

538.522:517.788

THE COMPUTATION OF THE INTEGRALS REQUIRED IN MUTUAL-INDUCTANCE CALCULATIONS

By N. F. ASTBURY, M.A., Electricity Department, National Physical Laboratory

Communicated by Dr L. Hartshorn, August 21, 1934. Read in title, November 16, 1934.

ABSTRACT. A method, based on the reduction of classical formulae, whereby the numerical computation of complete elliptic integrals of the third kind can be considerably simplified in certain cases is suggested. A particular case considered in detail is that which arises in the calculation of the mutual inductance of a helix and a coaxial circle. A considerably simplified formula, well adapted for numerical working, is deduced, together with an extremely simple result for the case when the helix and circle have the same radius.

---

§1. INTRODUCTION

The need for non-laborious methods of computing elliptic integrals arises in several physical problems. L. V. King[1] has given a very detailed account of the various devices by which elliptic integrals can be reduced to forms which admit of easy numerical computation, and, in addition to many classical formulae, he has given new series suitable for the direct numerical computation of complete elliptic integrals of the third kind. In certain cases, however, which are of interest in the calculation of coefficients of mutual inductance, it appeared to the present writer that much labour could be saved in the computation of the integrals of the third kind by the direct reduction of classical formulae to simpler forms. A new expression is derived which is well adapted for numerical working, and its application to Viriamu Jones' formula[2] for the mutual inductance of a helix and a coaxial circle is considered.

§2. REDUCTION OF CLASSICAL FORMULAE

$F(\phi, k)$, $E(\phi, k)$, (ϕ, k, n), k

Denoting by $F(\phi, k)$, $E(\phi, k)$ and $\Pi(\phi, k, n)$ respectively the incomplete elliptic integrals of the first, second and third kinds respectively to modulus k, we have

$$\left.\begin{aligned} F(\phi, k) &= \int_0^\phi d\phi/\Delta \\ E(\phi, k) &= \int_0^\phi \Delta\, d\phi \\ \Pi(\phi, k, n) &= \int_0^\phi d\phi/\{(1+n\sin^2\phi)\,\Delta\} \end{aligned}\right\} \qquad \text{......(1)},$$

Δ

where

$$\Delta = \sqrt{(1-k^2\sin^2\phi)}.$$

The angle ϕ is the *amplitude*, and n is the *parameter* of the third integral. The *complete integrals*, which are usually denoted by the letters K, E and Π respectively, are obtained by putting ϕ equal to $\pi/2$.

Methods for the evaluation of these integrals based on transformation of the moduli and amplitudes were discovered independently by Landen and by Gauss. Each method depends on the formation of a *scale of arithmetico-geometrical means*. Such a scale is formed as follows. Start with two positive numbers a_0, b_0, such that $a_0 > b_0$, and form successively the following quantities:

$$c_0 = \sqrt{(a_0^2 - b_0^2)}$$

$$\left.\begin{array}{lll} a_1 = \tfrac{1}{2}(a_0 + b_0) & b_1 = \sqrt{(a_0 b_0)} & c_1 = \tfrac{1}{2}(a_0 - b_0) \\ a_2 = \tfrac{1}{2}(a_1 + b_1) & b_2 = \sqrt{(a_1 b_1)} & c_2 = \tfrac{1}{2}(a_1 - b_1) \\ \cdots\cdots & \cdots\cdots & \cdots\cdots \\ a_n = \tfrac{1}{2}(a_{n-1} + b_{n-1}) & b_n = \sqrt{(a_{n-1} . b_{n-1})} & c_n = \tfrac{1}{2}(a_{n-1} - b_{n-1}) \end{array}\right\} \quad \ldots(2).$$

The a's and b's rapidly tend to the same limit, which we may denote by $a_n = M(a_0, b_0)$, while the c's become zero in the limit. It is easy to see that, if p is any number, then

$$p.M(a_0, b_0) = M(pa_0, pb_0) \qquad \ldots\ldots(3).$$

Landen's transformation is equivalent to forming the arithmetico-geometrical mean scale $a_0 = 1$, $b_0 = \sqrt{(1-k^2)}$, $c_0 = k$, together with a trigonometrical recurrence formula for the amplitudes (starting with $\phi_0 = \phi$)

$$\tan(\phi_{r+1} - \phi_r) = (b_r/a_r)\tan\phi_r \qquad \ldots\ldots(4),$$

and just as the a's and b's tend to a limit, so the ϕ's derived from equation (4) continually approach a value which is equal to 2^r multiplied by a definite magnitude. In other words, $\phi_r/2^r$ approaches a limit which we may denote by Φ.

We then have the following results[3] for incomplete integrals of the first and second kinds:

$$F(\phi, k) = \Phi/a_n \qquad \ldots\ldots(5),$$

$$E(\phi, k) = (E/K)\,F(\phi, k) + \sum_{r=1} c_r \sin\phi_r \qquad \ldots\ldots(6),$$

and for the complete integrals of the first and second kinds

$$K = \pi/2a_n \qquad \ldots\ldots(7),$$

$$(K-E)/K = \tfrac{1}{2}\sum_{r=0} 2^r c_r^2 \qquad \ldots\ldots(8).$$

We now proceed to apply these results to the reduction of the classical formulae for complete elliptic integrals of the third kind. Legendre showed that, in general, such an integral could be expressed in terms of complete and incomplete integrals of the first and second kinds, the result depending on the form of the parameter n of the third integral. In what follows we shall be concerned with the case when n is negative and particularly when n is such that $1 > |n| > k^2$, although in passing it is worth while noting, on account of its simplicity, the case in which $|n| < k^2$.

In this latter case, Legendre writes $n = -k^2 \sin^2\theta$ and deduces the result[4]

$$\Pi - K = \frac{\tan\theta}{\sqrt{(1 - k^2\sin^2\theta)}}\,[K.E(\theta, k) - E.F(\theta, k)] \qquad \ldots\ldots(9).$$

Forming the arithmetico-geometrical mean scale $a_0=1$, $b_0=\sqrt{(1-k^2)}$, $c_0=k$, using the recurrence formula (4), and applying results (5)–(8), we can easily throw this equation into the very simple form

$$\Pi-K=\frac{\pi\tan\theta}{2a_n\sqrt{(1-k^2\sin^2\theta)}}\sum_{r=1}c_r\sin\phi_r \qquad \text{......(10)},$$

which is very suitable for numerical work. If we work to 8-figure accuracy, the arithmetico-geometrical mean scale converges in four terms, while only two or three terms are required in the summation.

Turning now to the former case, in which we are directly interested from the point of view of mutual-inductance calculation, we note that Legendre writes

$$n=-1+k'^2\sin^2\phi_0' \qquad \text{......(11)},$$

k' in which $k'=\sqrt{(1-k^2)}$ and is called the *complementary modulus*. He derives the result[5]

$$\frac{k'^2\sin\phi_0'\cos\phi_0'}{\Delta(k'\,\phi_0')}(\Pi-K)=\pi/2+(K-E)\,F(\phi_0'\,k')-K.E(\phi_0'\,k') \qquad \text{...(12)}.$$

Form now the arithmetico-geometrical mean scale

$$a_0=1,\quad b_0=k',\quad c_0=\sqrt{(a_0^2-b_0^2)}=k$$

and its *complementary scale*

$$a_0'=1,\quad b_0'=c_0=k,\quad c_0'=k',$$

and in association with the complementary scale use the recurrence formula (4) starting with ϕ_0' as given by equation (11). Denoting by dashed letters all quantities related to the complementary scale and recurrence formula, we may, using results (5) to (8), write the right-hand side of equation (12) as

Φ'

$$\pi\left[\tfrac{1}{2}+\left(\tfrac{1}{2}\sum_{r=0}2^r c_r^2+\tfrac{1}{2}\sum_{r=0}2^r c_r'^2-1\right)\Phi'/(2a_n a_n')-(1/2a_n)\sum_{r=1}c_r'\sin\phi_r'\right] \qquad \text{......(13)}.$$

K', E' Denoting by dashed letters, K', E', complete integrals of the first and second kinds respectively to modulus k', we have[6]

$$EK'+E'K-KK'=\pi/2 \qquad \text{......(14)},$$

which may be written

$$(K-E)/K+(K'-E')/K'-1=-\pi/2KK' \qquad \text{......(15)},$$

which, by equations (7) and (8), becomes

$$\tfrac{1}{2}\sum_{r=0}2^r c_r^2+\tfrac{1}{2}\sum_{r=0}2^r c_r'^2-1=-2a_n a_n'/\pi \qquad \text{......(16)},$$

enabling us to substitute for the coefficient of Φ' in equation (13), which now becomes

$$\pi\left[\tfrac{1}{2}-\Phi'/\pi-\frac{1}{2a_n}\sum_{r=1}c_r'\sin\phi_r'\right] \qquad \text{......(17)}.$$

Finally, if we denote by Ψ the complement, *in degrees*, of Φ', equation (12) reduces to the form

$$(\Pi - K)/\pi = \{\Delta\,(k'\,\phi_0')/(2k'^2 \sin\phi_0' \cos\phi_0')\}\left\{\Psi/90 - \frac{1}{a_n}\sum_{r=1} c_r' \sin\phi_r'\right\}$$

$$= \{\sqrt{(-n)}/(k'^2 \sin 2\phi_0')\}\left\{\Psi/90 - \frac{1}{a_n}\sum_{r=1} c_r' \sin\phi_r'\right\} \qquad \text{......(18).}$$

This formula is very convenient for numerical work, comparing favorably with King's direct formula

$$\Pi/\pi = \frac{\cos(2\psi_1 - \psi_0)}{a_n k'^2 \sin 2\psi_0}\left\{a_n \sin\phi_n + 2\sum_{r=0}(2^{r+1}-1)\,c_{r+2}\,\frac{\tan(2\psi_{r+2}-\psi_{r+1})}{\cos(2\psi_{r+3}-\psi_{r+2})}\right\} \qquad \text{......(19),}$$

in which the arithmetico-geometrical mean scale $a_0 = 1$, $b_0 = k'$, $c_0 = k$ is used and the ψ's are derived from the recurrence formula

$$\sin(2\psi_{r+1} - \psi_r) = (b_r/a_r)\sin\psi_r \qquad \text{......(19.1),}$$

ψ_0 being given by

$$n = -k^2/(1 - k'^2 \sin^2\psi_0) \qquad \text{......(19.2).}$$

The process for numerical computation by formula (18) may now be briefly recapitulated as follows: (i) Construct the arithmetico-geometrical mean scale $a_0 = 1$, $b_0 = k'$, $c_0 = k$ and its complementary scale $a_0' = 1$, $b_0' = k$, $c_0' = k'$; (ii) derive Φ' or $\phi_n'/2^n$ from the recurrence formula $\tan(\phi'_{r+1} - \phi_r') = (b_r'/a_r')\tan\phi_r'$ starting with ϕ_0' as given by $n = -1 + k'^2 \sin^2\phi_0'$, and finally finding the complement Ψ of Φ'; (iii) effect the summation $\sum_{r=1} c_r' \sin\phi_r'$.

In general, when we are working to 8-figure accuracy, the first arithmetico-geometrical mean scale will converge in four terms and the complementary scale in two or three. The amplitude scale and the summation will converge in two or three terms, the third term in the latter having perhaps only two or three significant figures.

The advantages of formula (18) lie in its comparatively simple form, in the considerably reduced work required with 8-figure trigonometrical tables (the use of which rapidly becomes tedious) and in the very simple summation which is required. These more than outweigh the work involved in the formation of a second arithmetico-geometrical mean scale. In the numerical example given below only nine references to trigonometrical tables were required, as against seventeen when King's formula (19) was used, the whole computation occupying less than half the time.

The following figures, taken from an actual practical case, illustrate the use of formula (18):

$k^2 = 0{\cdot}8282\,6099$, $\quad k'^2 = 0{\cdot}1717\,3907$,

$n = -0{\cdot}9653\,3002$, $\quad 1 + n = 0{\cdot}0346\,6998$.

Arithmetico-geometrical mean scale: $a_0 = 1$, $b_0 = k'$, $c_0 = k$.

r	a	b	c
0	1·0000 0000	0·4144 1412	0·9100 8845
1	0·7072 0706	0·6437 5005	0·2927 9294
2	0·6754 7856	0·6747 3297	0·0317 2850
3	0·6751 0576	0·6751 0565	0·0003 7280
4	0·6751 0570	0·6751 0570	0·0000 0006

Arithmetico-geometrical mean scale: $a_0'=1$, $b_0'=k$, $c_0'=k'$.
Amplitude scale: $\Phi'=\phi_n'/2^n=(205^\circ\ 6'\ 39{\cdot}5'')/2^3=25^\circ\ 38'\ 19{\cdot}94''$.

r	a'	b'	c'	ϕ'
0	1·0000 0000	0·9100 8845	0·4144 1412	26° 41′ 57·06″
1	0·9550 4422	0·9539 8556	0·0449 5578	51° 17′ 35·71″
2	0·9545 1489	0·9545 1474	0·0005 2933	102° 33′ 19·77″
3	0·9545 1482	0·9545 1482	0·0000 0008	205° 6′ 39·54″

Hence $\quad \Psi=64^\circ\ 21'\ 40{\cdot}06''$ and $\Psi/90=0{\cdot}7151\ 2364$,

$$\frac{1}{a_n}\sum_{r=1} c_r'\sin\phi_r'=\frac{0{\cdot}0350\ 8155+0{\cdot}0005\ 1667-0{\cdot}0000\ 0003}{0{\cdot}6751\ 0570}$$

$$=0{\cdot}0527\ 2980;$$

$$\therefore\ \Psi/90-\frac{1}{a_n}\sum_{r=1} c_r'\sin\phi_r'=0{\cdot}6623\ 9384.$$

Also $\quad \sqrt{(-n)}/(k'^2\sin 2\phi_0')=7{\cdot}1262\ 523,$

so that $\quad (\Pi-K)/\pi=0{\cdot}6623\ 9384\times 7{\cdot}1262\ 523=4{\cdot}7203\ 856.$

§3. APPLICATION TO MUTUAL INDUCTANCE CALCULATION

J. Viriamu Jones [2] has given the following formula for the mutual inductance M between a helix of N turns, length x and radius B, and a coaxial circle of radius A in the plane of one end:

$$M=2\pi^2 N(A+B)\left\{\frac{g}{k}\,\frac{K-E}{\pi}+\frac{kg'^2}{g}\cdot\frac{K-\Pi}{\pi}\right\} \qquad \text{......(20)},$$

in which the modulus k of the elliptic integrals is given by

$$k^2=4AB/\{(A+B)^2+x^2\} \qquad \text{......(20.1)},$$

and the parameter n of the third integral by

$$n=-g^2=-4AB/(A+B)^2 \qquad \text{......(20.2)}.$$

In addition $\quad k^2+k'^2=1=g^2+g'^2 \qquad$ ......(20.3).

Let us now write

$$\left.\begin{aligned}R_1^2&=(A+B)^2+x^2\\ R_2^2&=(A-B)^2+x^2\end{aligned}\right\} \qquad \text{......(20.4)},$$

so that $\quad k'=R_2/R_1.$

It is easy to see in this case that equation (18) reduces to the simple form

$$(\Pi-K)/\pi=\frac{R_1}{2g'x}\left\{\frac{\Psi}{90}-\frac{1}{a_n}\sum_{r=1} c_r'\sin\phi_r'\right\} \qquad \text{......(21)}.$$

We note now that $(A+B)\,g\,k=R_1$. Bearing in mind the homogeneity relation (3), we may now form the arithmetico-geometrical mean scales

$$a_0=R_1,\qquad b_0=R_2,\qquad c_0=\sqrt{(R_1^2-R_2^2)}=\sqrt{(4AB)},$$

and $\quad a_0'=R_1,\qquad b_0'=\sqrt{(4AB)},\qquad c_0'=R_2,$

together with the recurrence formula (for deriving Ψ')

$$\tan(\phi'_{r+1}-\phi_r')=(b_r'/a_r')\tan\phi_r',$$

starting with $\phi_0'=\arc\sin(g'/k')$.

Denoting by a_n the limit to which the first scale converges, and by S_1 and S_2 respectively the summations $\sum\limits_{r=0} 2^r c_r^2$ and $\sum\limits_{r=1} c_r' \sin\phi_r'$, and applying results (8), (18) and (21) to (20), we get

$$M=2\pi^2 N\left[\frac{S_1}{4a_n}-\frac{A^2-B^2}{2x}\left\{\frac{\Psi'}{90}-\frac{S_2}{a_n}\right\}\right]$$

$$=\pi^2 N\left[\frac{1}{2}\frac{S_1}{a_n}-\frac{A^2-B^2}{x}\left\{\frac{\Psi'}{90}-\frac{S_2}{a_n}\right\}\right] \qquad \text{......(22)},$$

which is a form well adapted to numerical working.

The three dimensions A, B and x required for the problem being given, the process of calculation may be summarized as follows: (i) Find R_1 or $\sqrt{\{(A+B)^2+x^2\}}$ and R_2 or $\sqrt{\{(A-B)^2+x^2\}}$; (ii) construct the arithmetico-geometrical mean scales $a_0=R_1$, $b_0=R_2$, $c_0=\sqrt{(4AB)}$ and $a_0'=R_1$, $b_0'=\sqrt{(4AB)}$, $c_0'=R_2$; (iii) from the first scale derive the limit a_n and from the second scale and recurrence formula (4) derive Φ' and Ψ' starting with $\phi_0'=\arc\sin(g'/k')=\arc\sin\{R_1(A-B)/R_2(A+B)\}$; (iv) effect the summations S_1 and S_2.

Finally, it is worth noting the case when $A=B$. Equation (22) then reduces to the very simple form

$$M=\pi^2 N S_1/2a_n \qquad \text{......(23)}.$$

This formula, which is analogous to King's formula(7) for coaxial circles, gives the mutual inductance of a helix of N turns and a single turn of the same radius in the plane of one end.

It should be mentioned that Campbell(8), Smith(9) and Dye(10) all appear to have used Legendre's result (12) in the evaluation of Jones's formula, but the simplified form (22) derived above does not seem to have been known.

§ 4. ACKNOWLEDGMENTS

The writer is obliged to Dr L. Hartshorn and Mr T. Smith, F.R.S. for their criticism of the MS.

REFERENCES

(1) L. V. King. *On the Direct Numerical Calculation of Elliptic Functions and Integrals* (C.U.P. 1924).
(2) J. V. Jones. *Proc. Roy. Soc.* A, **63**, 198 (1898).
(3) King. *Loc. cit.* p. 8.
(4) Legendre. *Fonctions Elliptiques*, **1**, 148.
(5) Legendre. *Ibid.* **1**, 151.
(6) Cayley. *Elliptic Functions* (2nd edition), p. 48.
(7) L. V. King. *Proc. Roy. Soc.* A, **100**, 63 (1921).
(8) A. Campbell. *Proc. Roy. Soc.* A, **87**, 391 (1912).
(9) F. E. Smith. *Philos. Trans.* A, **214**, 211 (1912).
(10) D. W. Dye. *Proc. Roy. Soc.* A, **101**, 315 (1922).

Further accounts of the use of arithmetico-geometrical mean series in inductance calculations will be found in papers by L. V. King and by F. W. Grover, *Phil. Mag.* Series 7, **15**, 1097 *et seq.* (June 1933).

519.281

ON OBSERVATIONS OF POINTS CONNECTED BY A LINEAR RELATION

By Miss BERYL M. DENT, M.Sc., Research Department, Metropolitan-Vickers Electrical Co. Ltd.

Communicated by H. R. Hassé, July 10, 1934. *Read in title, November* 16, 1934.

ABSTRACT. The problem of drawing the best straight line through a set of observed points is solved by a method shorter than those previously published. It is essential for the complete solution of the problem to obtain the most probable value of the ratio of the precision constants of the two observed sets of quantities and a new method is given for finding this ratio. Expressions for the errors in the position and inclination of the line are derived and a numerical example is added.

§ 1. INTRODUCTION

The problem of drawing the best straight line through a set of observed points is usually treated in text books on the assumption (often tacit) that only one of the coordinates is liable to error. The more general case in which both coordinates are liable to error has been dealt with by Pearson*, Stewart† and Uhler‡. Pearson lays down an arbitrary criterion for a good fit, Stewart gives the solution for the case in which the precision constants for the two coordinates are equal, and Uhler obtains the solution when the two precision constants are not equal. The general solution is however dependent on a knowledge of the ratio of the precision constants, and no indication is given of any method of estimating the values of the precision constants from the given data if these consist, as they usually do, of a series of single observations of different points. Moreover, no consideration of the errors in the position and inclination of the line are given. In this paper both these matters are dealt with. The solution of the problem when the line is constrained to pass through a given point follows without further analysis.

A paper by W. R. Cook§ gives a solution of the problem of curve-fitting by means of least squares, including a consideration of the errors and a determination of the most probable value of the ratio of the precision constants. His method, however, is only applicable to curves represented by equations of the second and higher degrees, and not to a straight line.

* Pearson, K., *Phil. Mag.* 6, **2**, 559 (1901).
† Stewart, R. Meldrum, *Phil. Mag.* 6, **40**, 217 (1920).
‡ Uhler, H. S., *J. opt. Soc. Amer.* **7**, 1043 (1923).
§ Cook, W. R., *Phil. Mag.* 7, **12**, 1025 (1931).

§ 2. STATEMENT OF THE PROBLEM

Given a set of S observed points (X_1, Y_1), $(X_2, Y_2) \ldots (X_s, Y_s)$, whose true positions are known from theoretical considerations to lie on a straight line, the problem is to find the best straight line through them when both the X's and the Y's are liable to error. $S, (X_r, Y_r)$

It will be assumed, to start with, that repeated observations of any one point may be made, and that it is possible to keep either coordinate constant and make repeated observations of the other. The repeated observations allow error laws for both coordinates to be formulated. It will be assumed that the error laws for all the X's are the same and that those for all the Y's are the same, but different from those for the X's. In a later paragraph a method of solution will be given for cases where these assumptions are not justified.

We shall now solve the problem of finding the straight line on which the points (x_r, y_r) lie, when the x_r and y_r are chosen so as to make it most probable that the points (x_r, y_r) are the true positions of the observed points (X_r, Y_r). The points (x_r, y_r) will be referred to as the best* points. (x_r, y_r)

§ 3. SOLUTION. GENERAL CASE

Let $\Phi(\Delta)$ represent the frequency distribution of errors of observation in the X's and $\Phi'(\Delta)$ represent the frequency distribution of errors of observation in the Y's. $\Phi(\Delta)$ and $\Phi'(\Delta)$ are assumed to be normalized, i.e. $\Phi(\Delta)$ $\Phi'(\Delta)$

$$\int_{-\infty}^{+\infty} \Phi(\Delta)\, d\Delta = \int_{-\infty}^{+\infty} \Phi'(\Delta)\, d\Delta = 1.$$

Then the probability that the best position of the point observed as (X_r, Y_r) is (x_r, y_r) is equal to

$$\Phi(x_r - X_r)\, \Phi'(y_r - Y_r)\, \delta x_r \delta y_r$$

and the probability that the best positions of all the observed points are

$$(x_1, y_1), (x_2, y_2) \ldots (x_s, y_s)$$

* The word *best* has been used throughout this paper as equivalent to *most probable*. Exception may reasonably be taken to the use of the word *true* in this connection. Strictly speaking, we cannot hope to find the true positions of the observed points, we can only make a guess at their true positions. In other words, we can find those positions which, when all the given data are taken into consideration, are most probable for the observed points. Obviously, if further data come into our possession, such as a detection of systematic error, the best values found when these data also are taken into consideration will be different from those previously found. We have then made a better guess at the true values.

The admission that we cannot hope to find the true positions of the observed points, but only their most probable positions, enables us to avoid the bugbear of inverse probability.

Let the frequency-distribution diagram of repeated observations on a quantity X_r have a maximum at X_r^m and let the errors be the deviations from X_r^m. Then X_r^m is the *best* or most probable value of X_r. If $\Phi(X_r^m - X_r)$ represents the frequency-distribution of the errors, the probability that X_r^m shall be observed as lying between X_r and $X_r + \delta X_r$ is $\Phi(X_r^m - X_r)\, \delta X_r$ and $\Phi(X_r^m - X_r)\, \delta X_r^m$ is the probability that the point observed as X_r shall have a *best*, or most probable, value lying between X_r^m and $X_r^m + \delta X_r^m$. This latter statement is rigorously correct, and no assumption such as "*a priori*, all values of X_r^m are equally likely" is necessary. X_r^m

is equal to

P $$\prod_r \Phi(x_r - X_r)\, \Phi'(y_r - Y_r)\, \delta x_r \delta y_r = P,$$

say.

P is therefore a function of the $2s$ unknown quantities x_r, y_r. By making it a maximum, subject to the condition that the points (x_r, y_r) are collinear, we shall obtain the parameters of the required line, and also, if we desire them, the values of all the coordinates x_r and y_r.

§ 4. SOLUTION. GAUSSIAN ERROR FUNCTIONS

h Let $$\Phi(\Delta) = \frac{h}{\sqrt{\pi}} e^{-h^2 \Delta^2}$$

h' and $$\Phi'(\Delta) = \frac{h'}{\sqrt{\pi}} e^{-h'^2 \Delta^2},$$

then we have to make the expression

$$\prod_r \frac{hh'}{\pi} \exp\left[-\{h^2 (x_r - X_r)^2 + h'^2 (y_r - Y_r)^2\}\right] \delta x_r \delta y_r$$

a maximum, subject to the conditions

m, c $$y_r = mx_r + c, \qquad r = 1, 2 \ldots s.$$

We must therefore make

$$h^2 \sum_{r=1}^{s} (x_r - X_r)^2 + h'^2 \sum_{r=1}^{s} (y_r - Y_r)^2$$

a minimum, subject to the conditions

$$y_r = mx_r + c.$$

Using the method of Lagrangian multipliers, we form the expression

$(x_1 \ldots \lambda_r)$ $F(x_1, x_2 \ldots x_s, y_1, y_2 \ldots y_s, m, c, \lambda_r)$

λ_0, λ_r $$\equiv \lambda_0 \left[h^2 \Sigma (x_r - X_r)^2 + h'^2 \Sigma (y_r - Y_r)^2\right] + \Sigma \lambda_r (y_r - mx_r - c) = 0,$$

$$\frac{\partial F}{\partial x_r} = \lambda_0\, 2h^2 (x_r - X_r) - m\lambda_r = 0, \qquad r = 1, 2 \ldots s \qquad \ldots\ldots(1),$$

$$\frac{\partial F}{\partial y_r} = \lambda_0\, 2h'^2 (y_r - Y_r) + \lambda_r = 0, \qquad r = 1, 2 \ldots s \qquad \ldots\ldots(2),$$

$$\frac{\partial F}{\partial m} = -\Sigma \lambda_r x_r = 0 \qquad \ldots\ldots(3),$$

$$\frac{\partial F}{\partial c} = -\Sigma \lambda_r = 0 \qquad \ldots\ldots(4),$$

$$\frac{\partial F}{\partial \lambda_r} = y_r - mx_r - c = 0, \qquad r = 1, 2 \ldots s \qquad \ldots\ldots(5).$$

These $3s + 2$ equations suffice to determine the $3s + 2$ unknown quantities

$$x_r,\ y_r,\ m,\ c,\ \lambda_r.$$

By summing equation (1) over all values of r and using (4), we obtain

$$\Sigma\,(x_r - X_r) = 0,$$

and similarly

$$\Sigma\,(y_r - Y_r) = 0,$$

i.e. the centroid of the observed points coincides with the centroid of the best points. Hence the required line passes through the centroid $(\bar{X}, \bar{Y})$ of the observed points. Change the origin to the point $(\bar{X}, \bar{Y})$ and let the new coordinates be denoted by dashed letters, so that $x_r' = x_r - \bar{X}$, and so on. $(\bar{X}, \bar{Y})$ x_r', y_r'

Then equations (1) to (5) are replaced by

$$\lambda_0\, 2h^2\,(x_r' - X_r') - m\lambda_r = 0, \qquad r = 1, 2 \ldots s \qquad \ldots\ldots(6),$$ X_r'

$$\lambda_0\, 2h'^2\,(y_r' - Y_r') + \lambda_r = 0, \qquad r = 1, 2 \ldots s \qquad \ldots\ldots(7),$$ Y_r'

$$\Sigma \lambda_r x_r' = 0 \qquad \ldots\ldots(8),$$

$$y_r' - m x_r' = 0, \qquad r = 1, 2 \ldots s \qquad \ldots\ldots(9).$$

Eliminating y_r' from equation (7) by means of (9), we have

$$2\lambda_0\, h'^2\,(m x_r' - Y_r') + \lambda_r = 0 \qquad \ldots\ldots(10).$$

From equation (6)

$$\frac{\lambda_r}{\lambda_0} = \frac{1}{m}\, 2h^2\,(x_r' - X_r') \qquad \ldots\ldots(11).$$

Substituting from equation (11) in (10) and (8), we have

$$h'^2\,(m x_r' - Y_r') + (h^2/m)\,(x_r' - X_r') = 0 \qquad \ldots\ldots(12)$$

and

$$\Sigma\,(x_r' - X_r')\, x_r' = 0 \qquad \ldots\ldots(13).$$

From equation (12)

$$x_r' = \frac{m h'^2 Y_r' + h^2 X_r'}{m^2 h'^2 + h^2} \qquad \ldots\ldots(14).$$

On substitution in equation (13)

$$\Sigma \left(\frac{m h'^2 Y_r' + h^2 X_r'}{m^2 h'^2 + h^2}\right)^2 - \Sigma \left(\frac{m h'^2 Y_r' + h^2 X_r'}{m^2 h'^2 + h^2}\right) X_r' = 0$$

or

$$m^2 h'^2 \Sigma X_r' Y_r' - m\,(h'^2 \Sigma Y_r'^2 - h^2 \Sigma X_r'^2) - h^2 \Sigma X_r' Y_r' = 0 \qquad \ldots\ldots(15).$$

Write $h/h' = k$. Then equation (15) becomes k

$$m^2 \Sigma X_r' Y_r' - m\,(\Sigma Y_r'^2 - k^2\, \Sigma X_r'^2) - k^2 \Sigma X_r' Y_r' = 0 \qquad \ldots\ldots(16).$$

A convenient method of solving this quadratic is as follows:

$$\frac{2mk}{k^2 - m^2} = \frac{2k \Sigma X_r' Y_r'}{k^2 \Sigma X_r'^2 - \Sigma Y_r'^2},$$

from equation (16). Now write

$$\tan\phi = (1/k)\tan\theta = m/k \qquad \ldots\ldots(17).$$ ϕ, θ

Then

$$\tan 2\phi = \frac{2m/k}{1 - m^2/k^2} = \frac{2mk}{k^2 - m^2}.$$

Therefore

$$\tan 2\phi = 2k\, \Sigma\, X_r' Y_r' / (k^2\, \Sigma\, X_r'^2 - \Sigma\, Y_r'^2) \qquad \ldots\ldots(18).$$

ϕ may therefore be found first from equation (18) and then m from equation (17). The position and inclination of the required line are therefore completely determined. The coordinates x_r', y_r' may also be found from equations (14) and (9). Of the two solutions for m, the correct one is that which makes

$$h^2 \Sigma (x_r' - X_r')^2 + h'^2 \Sigma (y_r' - Y_r')^2$$

a minimum.

If the error laws have been determined by repeated observations of one of the points (X_r, Y_r) the most probable position (X_r^m, Y_r^m) of this point will have been found. We therefore choose this point as origin instead of the centroid of the observed points, and proceed as from equation (6). The value of m is still given by equations (17) and (18), but X', Y' now refer to (X_r^m, Y_r^m) as origin.

§ 5. SPECIAL CASES

(a) *The case in which $h = h'$.* When the precision constants for X and Y are equal, i.e. when $h = h'$, the expressions are considerably simplified. Equations (17) and (18) give

$$\tan\phi = \tan\theta = m$$

and

$$\tan 2\theta = 2(\Sigma X_r' Y_r')/(\Sigma X_r'^2 - \Sigma Y_r'^2) \qquad \text{......(19).}$$

The ambiguity as to $\pi/2$ is resolved by making $\Sigma\{(x_r' - X_r')^2 + (y_r' - Y_r')^2\}$ a minimum.

From equations (6) and (7) we have, when $h = h'$,*

$$\frac{y_r' - Y_r'}{x_r' - X_r'} = -\frac{1}{m},$$

so that the best points are the feet of the perpendiculars from the observed points to the required line, and the required line is that which makes the sum of the squares of the perpendiculars to it from the observed points a minimum.

(b) *The case in which $h \to \infty$, $h' \to \infty$.* If the observed values of X_r may be regarded as free from error, i.e. if $h \to \infty$, equation (15) becomes

$$m = (\Sigma X_r' Y_r')/(\Sigma X_r'^2) \qquad \text{......(20).}$$

Similarly, if $h' \to \infty$,

$$m = (\Sigma Y_r'^2)/(\Sigma X_r' Y_r') \qquad \text{......(21).}$$

§ 6. CONSIDERATION OF ERRORS

$$\bar{X} = (\Sigma X_r)/s, \qquad \bar{Y} = (\Sigma Y_r)/s.$$

Following Eddington†, we shall express the errors in the centroid of the observed points and the inclination of the required line in terms of "range per risk."

μ_r (a) *Centroid of observed points.* The probability of an error μ_r in X_r is

$$\frac{h}{\sqrt{\pi}} e^{-h^2\mu_r^2} d\mu_r.$$

* In general $\dfrac{y_r' - Y_r'}{x_r' - X_r'} = \dfrac{h^2}{h'^2}\left(-\dfrac{1}{m}\right)$, so that all the lines joining the best to the observed points are parallel. This property holds good when the true points lie on any curve or surface, and it has previously been noted by Deming, *Phil. Mag.* 11, 146, 1931, and Uhler, *loc. cit.*

† *Proc. Phys. Soc. Lond.* 45, 271 (1933).

Therefore the probability of an error $s^{-1}\Sigma\mu_r$ in $\bar{X}$ is

$$(h^s/\pi^{s/2})\exp(-h^2\Sigma\mu_r^2)\,d\mu_1\ldots d\mu_s \quad \ldots\ldots(22),$$

and the probability that the error in $\bar{X}$ has a value between ϵ_1 and ϵ_2 is the integral of equation (22) taken over the field of integration for which $s^{-1}\Sigma\mu_r$ lies between ϵ_1 and ϵ_2. ϵ_1, ϵ_2

This probability is therefore*

$$=\frac{1}{2\pi}\int_{-\infty}^{+\infty}\int_{\epsilon_1}^{\epsilon_2}\int_{-\infty}^{+\infty}\cdots\left\{\exp i\theta\left(\tau-\frac{1}{s}\Sigma\mu_r\right)\right\}(h^s/\pi^{s/2})\{\exp(-h^2\Sigma\mu_r^2)\}\times d\mu_1\ldots d\mu_s\,d\tau\,d\theta,$$

and on performing the integrations we find that the probability that the error in $\bar{X}$ lies between μ and $\mu+d\mu$ is $\psi(\mu)\,d\mu$, where $\psi(\mu)$

$$\psi(\mu)=\frac{h}{\sqrt{\pi}}\sqrt{s}\,.\,e^{-\mu^2h^2s} \quad \ldots\ldots(23).$$

Similarly the probability that the error in $\bar{Y}$ lies between μ and $\mu+d\mu$ is $\psi'(\mu)\,d\mu$, where $\psi'(\mu)$

$$\psi'(\mu)=\frac{h'}{\sqrt{\pi}}\sqrt{s}\,.\,e^{-\mu^2h'^2s}.$$

Hence the probability that the modulus of the error in $\bar{X}$ is not greater than $\mu/\sqrt{2}$ is

$$2\int_0^{\mu/\sqrt{2}}\psi(\mu)\,d\mu \quad\text{or}\quad p,\text{ say}, \quad \ldots\ldots(24),$$ p

and the probability that the modulus of the error in $\bar{Y}$ is not greater than $\mu/\sqrt{2}$ is

$$2\int_0^{\mu/\sqrt{2}}\psi'(\mu)\,d\mu \quad\text{or}\quad p',\text{ say},$$ p'

so that the probability that the best position of the centroid shall not be farther than μ from $(\bar{X}, \bar{Y})$ is pp'. I therefore take a risk of 1 in $1/(1-pp')$ of being wrong by asserting that the best position of the centroid is not farther than μ from $(\bar{X}, \bar{Y})$.

(*b*) *Inclination of the required line*. The error in m can readily be deduced in any special case from the error in $\tan 2\phi$ by means of the equation $m=k\tan\phi$. If the error in X_r' be μ_r and the error in Y_r' be ν_r, then ν_r

$$\tan 2\phi=\frac{2k\Sigma(X_r'+\mu_r)(Y_r'+\nu_r)}{k^2\Sigma(X_r'+\mu_r)^2-\Sigma(Y_r'+\nu_r)^2}.$$

Expanding by Taylor's theorem, we find that the error in $\tan 2\phi$ is

$$\frac{2k\Sigma\mu_rY_r'}{(k^2\Sigma X_r'^2-\Sigma Y_r'^2)}-\frac{4k\Sigma X_r'Y_r'\Sigma\mu_rX_r'}{(k^2\Sigma X_r'^2-\Sigma Y_r'^2)^2}$$
$$+\frac{2k\Sigma\nu_rX_r'}{(k^2\Sigma X_r'^2-\Sigma Y_r'^2)}-\frac{4k\Sigma X_r'Y_r'\Sigma\nu_rY_r'}{(k^2\Sigma X_r'^2-\Sigma Y_r'^2)^2}.$$

Writing $A=2k/(k^2\Sigma X_r'^2-\Sigma Y_r'^2)$ and $B=4k\Sigma X_r'Y_r'/(k^2\Sigma X_r'^2-\Sigma Y_r'^2)^2$, we have that the error in $\tan 2\phi$ A, B

$$=A\Sigma\mu_rY_r'-k^2B\Sigma\mu_rX_r'+A\Sigma\nu_rX_r'+B\Sigma\nu_rY_r'.$$

* Whittaker and Robinson, *Calculus of Observations*, pp. 168 and 175 (1929).

Now the probability that there is an error μ_r in X_r' is

$$\frac{h}{\sqrt{\pi}} e^{-h^2 \mu_r^2} d\mu_r,$$

and the probability that errors μ_r and ν_r occur simultaneously in the values X_r' and Y_r' is

$$(hh'/\pi)^s \exp\left[-\{h^2 \Sigma \mu_r^2 + h'^2 \Sigma \nu_r^2\}\right].$$

The probability that the error in $\tan 2\phi$ has a value between ϵ_1 and ϵ_2 is the integral of this expression taken over the field of integration for which

$$A \Sigma \mu_r Y_r' - k^2 B \Sigma \mu_r X_r' + A \Sigma \nu_r X_r' + B \Sigma \nu_r Y_r'$$

lies between ϵ_1 and ϵ_2.

By the method of the preceding paragraph it can be shown that the probability

$\chi(\mu)$ that the error in $\tan 2\phi$ lies between μ and $\mu + d\mu$ is $\chi(\mu)\, d\mu$, where

$$\chi(\mu) = \frac{1}{\sqrt{\pi}} \frac{e^{-\mu^2/E^2}}{E} \qquad \text{......(25)},$$

in which

E $$E^2 = h^{-2} \Sigma (AY_r' - k^2 BX_r')^2 + h'^{-2} \Sigma (AX_r' + BY_r')^2$$
$$= h^{-2} \{(A^2 + k^2 B^2)(\Sigma Y_r'^2 + k^2 \Sigma X_r'^2)\}.$$

The probability that the modulus of the error in $\tan 2\phi$ is not greater than ϵ is therefore

q $$2 \int_0^\epsilon \chi(\mu)\, d\mu \quad \text{or} \quad q, \text{ say}, \qquad \text{......(26)}.$$

I therefore take a risk of 1 in 1 $(1-q)$ of being wrong by asserting that the best value of $\tan 2\phi$ does not differ by more than ϵ from the value

$$2k (\Sigma X_r' Y_r')/(k^2 \Sigma X_r'^2 - \Sigma Y_r'^2).$$

§ 7. MOST PROBABLE VALUES OF THE PRECISION CONSTANTS

The foregoing analysis may be applied directly to the problem provided that the precision constants for the X and Y coordinates can be determined. If this is not possible, we must try to use the given data to determine their most probable values.

In the expression

$$\prod_r (hh'/\pi) \exp\left[-\{h^2 (x_r - X_r)^2 + h'^2 (y_r - Y_r)^2\}\right] \delta x_r \delta y_r,$$

which is equal to the probability that the best positions of all the observed points are $(x_1, y_1) \ldots (x_s, y_s)$, h and h' must now be regarded as variables.

Substituting $y_r = mx_r + c$ in this expression, we write

$F'(x_1 \ldots c)$ $$F'(x_1 \ldots x_s, h, h', m, c) \equiv (hh'/\pi)^s \exp\left[-\{h^2 \Sigma (x_r - X_r)^2 + h'^2 \Sigma (y_r - Y_r)^2\}\right].$$

For maximum probability, F' is a maximum and therefore

$$\frac{\partial F'}{\partial x_r} = \frac{\partial F'}{\partial h} = \frac{\partial F'}{\partial h'} = \frac{\partial F'}{\partial m} = \frac{\partial F'}{\partial c} = 0.$$

These equations give

$$h^2(x_r - X_r) + h'^2(mx_r + c - Y_r)\,m = 0, \qquad r = 1, 2 \ldots s \qquad \ldots\ldots(27),$$

$$s = 2h^2 \Sigma (x_r - X_r)^2 \qquad \ldots\ldots(28),$$

$$s = 2h'^2 \Sigma (mx_r + c - Y_r)^2 \qquad \ldots\ldots(29),$$

$$\Sigma (mx_r + c - Y_r)\,x_r = 0 \qquad \ldots\ldots(30),$$

$$\Sigma (mx_r + c - Y_r) = 0 \qquad \ldots\ldots(31).$$

From equations (31) and (27) the centroid theorem follows as before.

Changing the origin to the centroid of the observed points and denoting the new coordinates by dashed letters, we therefore have the following equations to determine h, h', x_r', m:

$$h^2(x_r' - X_r') + h'^2(mx_r' - Y_r')\,m = 0, \qquad r = 1, 2 \ldots s \qquad \ldots\ldots(32),$$

$$s = 2h^2 \Sigma (x_r' - X_r')^2 \qquad \ldots\ldots(33),$$

$$s = 2h'^2 \Sigma (mx_r' - Y_r')^2 \qquad \ldots\ldots(34),$$

$$\Sigma x_r' (mx_r' - Y_r') = 0 \qquad \ldots\ldots(35).$$

From equation (32) by squaring:

$$h^4(x_r' - X_r')^2 = h'^4(mx_r' - Y_r')^2 m^2, \qquad r = 1, 2 \ldots s$$

and summing over all values of r, we have

$$\frac{h^4}{h'^4} = m^2 \frac{\Sigma (mx_r' - Y_r')^2}{\Sigma (x_r' - X_r')^2}.$$

From equations (33) and (34)

$$\frac{h^2}{h'^2} = \frac{\Sigma (mx_r' - Y_r')^2}{\Sigma (x_r' - X_r')^2}.$$

Hence

$$h^2/h'^2 = m^2 \qquad \ldots\ldots(36).$$

Substituting in equations (32) and (35) and solving for m and x_r', we have

$$x_r' = (mX_r' + Y_r')/2m \qquad \ldots\ldots(37)$$

and

$$m^2 = (\Sigma Y_r'^2)/(\Sigma X_r'^2) \qquad \ldots\ldots(38).$$

The most probable values of h and h' may now be determined separately from equations (33) and (34).

§ 8. ERRORS CORRESPONDING TO THE MOST PROBABLE VALUES OF h AND h'

(*a*) *Centroid of observed points.* The analysis of § 6 (*a*) is unaltered; we simply use the most probable values of h and h' in the expressions for $\psi(\mu)$ and $\psi'(\mu)$.

(*b*) *Inclination of the required line.* When k has its most probable value,

$$m = (\Sigma Y_r'^2)^{\frac{1}{2}} (\Sigma X_r'^2)^{-\frac{1}{2}}.$$

If μ_r and ν_r are the errors in X_r' and Y_r' respectively,

$$\begin{aligned} m &= \{\Sigma (Y_r' + \nu_r)^2\}^{\frac{1}{2}} \{\Sigma (X_r' + \mu_r)^2\}^{-\frac{1}{2}} \\ &= (\Sigma Y_r'^2)^{\frac{1}{2}} (\Sigma X_r'^2)^{-\frac{1}{2}} - (\Sigma Y_r'^2)^{\frac{1}{2}} (\Sigma X_r'^2)^{-\frac{3}{2}} \Sigma X_r' \mu_r \\ &\quad + (\Sigma Y_r'^2)^{-\frac{1}{2}} (\Sigma X_r'^2)^{-\frac{1}{2}} \Sigma Y_r' \nu_r, \text{ etc.} \end{aligned}$$

The error in m is therefore $A' \Sigma X_r' \mu_r + B' \Sigma Y_r' \nu_r$, where

A'

$$A' = -(\Sigma Y_r'^2)^{\frac{1}{2}} (\Sigma X_r'^2)^{-\frac{3}{2}},$$

B'

$$B' = (\Sigma Y_r'^2)^{-\frac{1}{2}} (\Sigma X_r'^2)^{-\frac{1}{2}}.$$

By the same method as before it may be deduced that the probability that the

$\chi'(\mu)$ error in m lies between μ and $\mu + d\mu$ is $\chi'(\mu)\, d\mu$, where

$$\chi'(\mu) = \frac{1}{\sqrt{\pi}} \frac{e^{-\mu^2/E'^2}}{E'} \qquad \text{......(39)},$$

E'

and

$$E'^2 = \frac{1}{h^2} \Sigma (A' X_r')^2 + \frac{1}{h'^2} \Sigma (B' Y_r')^2 = m^2 \left(\frac{1}{h^2} + \frac{1}{h'^2} \right).$$

The probability that the modulus of the error in $\tan\theta$ is not greater than ϵ is therefore equal to

$$2 \int_0^\epsilon \chi'(\mu)\, d\mu \quad \text{or } q', \text{ say} \qquad \text{......(40)}.$$

I therefore take a risk of 1 in $1/(1-q')$ of being wrong by asserting that the best value of m does not differ by more than ϵ from the value $(\Sigma Y_r'^2)^{\frac{1}{2}} (\Sigma X_r'^2)^{-\frac{1}{2}}$.

§ 9. NUMERICAL EXAMPLE

As an application let us take the four-point example quoted by Uhler*, Stewart† and Merriman‡.

The observed points are

$$X = 0{\cdot}4, \quad 0{\cdot}6, \quad 0{\cdot}8, \quad 0{\cdot}9;$$
$$Y = 0{\cdot}5, \quad 0{\cdot}8, \quad 1{\cdot}0, \quad 1{\cdot}2.$$

We find first $\bar{X}$ and $\bar{Y}$:

$$\bar{X} = 2{\cdot}7/4 = 0{\cdot}675;$$
$$\bar{Y} = 3{\cdot}5/4 = 0{\cdot}875.$$

The coordinates referred to the centroid as origin are:

$$X' = -0{\cdot}275, \quad -0{\cdot}075, \quad +0{\cdot}125, \quad +0{\cdot}225;$$
$$Y' = -0{\cdot}375, \quad -0{\cdot}075, \quad +0{\cdot}125, \quad +0{\cdot}325.$$

Since h and h' are not given, we must use equation (38) to find the most probable value of m:

$$m^2 = (\Sigma Y_r'^2)/(\Sigma X_r'^2) = 0{\cdot}2675/0{\cdot}1475 = 1{\cdot}813,$$
$$m = 1{\cdot}347, \quad \theta = \tan^{-1} m = 53^\circ\, 25'.$$

The line is shown in figure 1.

If we assume, as Stewart does, that the precision constants h and h' are equal, we have to use the formula (19), viz.

$$\tan 2\theta = \frac{2\Sigma X_r' Y_r'}{\Sigma X_r'^2 - \Sigma Y_r'^2} = \frac{2 \times 0{\cdot}1975}{0{\cdot}1475 - 0{\cdot}2675} = -3{\cdot}29.$$

Therefore $$2\theta = 106^\circ\, 54', \qquad \theta = 53^\circ\, 27'.$$

* Uhler, *J. opt. Soc. Amer.* **7**, 1057 (1923).

† Stewart, *Phil. Mag.* 6, **40**, 223 (1920).

‡ Merriman, *Textbook on the Method of Least Squares*. 1885 edn. p. 127. Art. 107.

The two values to be compared with Stewart's are

(1) $\tan\theta = 1{\cdot}349$,

(2) the intercept on the y-axis $= \overline{Y} - m\overline{X} = 0{\cdot}875 - 1{\cdot}349 \times 0{\cdot}675 = -0{\cdot}035$,

both of which are exactly equal to Stewart's second (final) approximation.

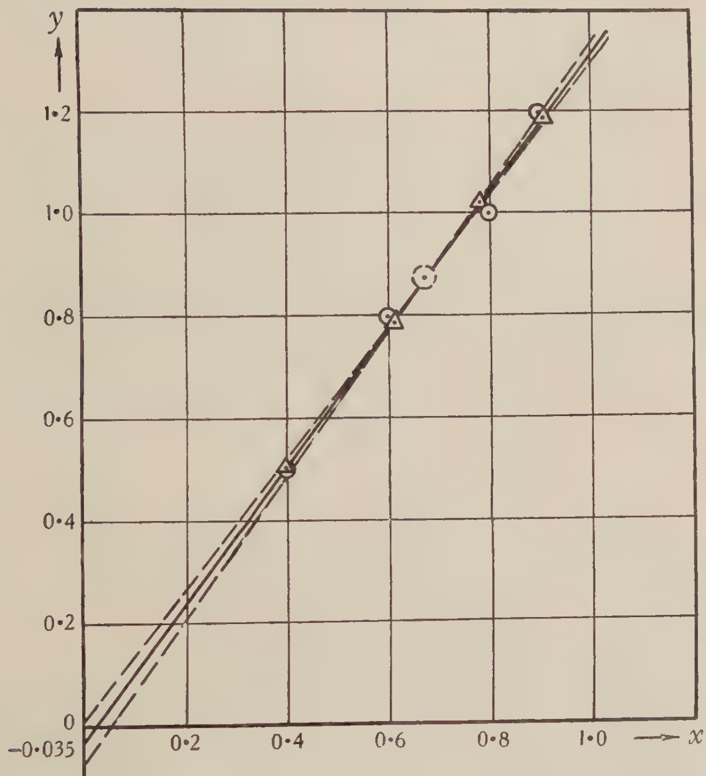


Figure 1. Straight line drawn through four observed points. The dotted circle and lines indicate the errors for a risk of 1 in 100. ⊙, observed. ▽, calculated.

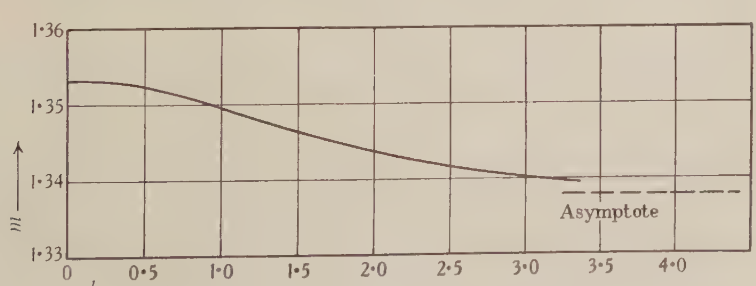


Figure 2. m as a function of k, equation (16)

We note that the two values of m obtained by making $k = 1$ or $1{\cdot}35$ differ by less than 1 per cent. It is now of interest to find the effect on m of allowing k to

range over its whole series of possible values, from 0 to ∞. The extreme values are found to be $m=1{\cdot}353$ corresponding to $k=0$, and $m=1{\cdot}338$ corresponding to $k=\infty$. The curve m against k is shown in figure 2. In this particular case, then, we can obtain the value of m correct to 1 per cent, without even knowing the ratio of the precision constants.

Errors. In order to find the errors we must determine the most probable values of h and h', using equations (33), (34) and (37)*. The work is set out in the following table ($m=1{\cdot}347$).

Table 1. Determination of the most probable values of h and h'

X'	$-0{\cdot}275$	$-0{\cdot}075$	$+0{\cdot}125$	$+0{\cdot}225$	Σ
mX'	$-0{\cdot}370$	$-0{\cdot}101$	$+0{\cdot}168$	$+0{\cdot}303$	—
Y'	$-0{\cdot}375$	$-0{\cdot}075$	$+0{\cdot}125$	$+0{\cdot}325$	—
$mX'+Y'$	$-0{\cdot}745$	$-0{\cdot}176$	$+0{\cdot}293$	$+0{\cdot}628$	—
$\frac{mX'+Y'}{2}=y'$	$-0{\cdot}372$	$-0{\cdot}088$	$+0{\cdot}146_5$	$+0{\cdot}314$	—
$\frac{mX'+Y'}{2m}=x'$	$-0{\cdot}277$	$-0{\cdot}065$	$+0{\cdot}109$	$+0{\cdot}233$	—
$(x_r'-X_r')$	$-0{\cdot}002$	$+0{\cdot}010$	$-0{\cdot}016$	$+0{\cdot}008$	—
$(x_r'-X_r')^2$	4×10^{-6}	100×10^{-6}	256×10^{-6}	64×10^{-6}	42×10^{-5}
$(y_r'-Y_r')$	$+0{\cdot}003$	$-0{\cdot}013$	$+0{\cdot}021_5$	$-0{\cdot}011$	—
$(y_r'-Y_r')^2$	9×10^{-6}	169×10^{-6}	462×10^{-6}	121×10^{-6}	76×10^{-5}

This table gives us, incidentally, the positions of the best points x_r', y_r'. These are shown in figure 1. From equation (33)

$$h^2=\frac{4}{2\Sigma\,(x_r'-X_r')^2}=\frac{2}{42\times10^{-5}}=4760, \qquad h=69{\cdot}0;$$

and from equation (34)

$$h'^2=\frac{4}{2\Sigma\,(y_r'-Y_r')^2}=\frac{2}{76\times10^{-5}}=2630, \qquad h'=51{\cdot}3.$$

As a check on the work we note that $h^2/h'^2=4{\cdot}760/2{\cdot}630=1{\cdot}811=(1{\cdot}346)^2$.

* The sums $\Sigma\,(x_r-X_r)^2$ and $\Sigma\,(y_r-Y_r)^2$, which are necessary to find h and h' separately, may be obtained very simply by the use of a planimeter. The method is due to C. F. Merriman (see the discussion following a paper by E. O. Waters on "Graphical Methods for Least-Square Problems," *A.S.M.E. Trans.* **51**, 201 (1929)).

The slope of all the lines joining the observed points to the best points when h and h' have their most probable values is $-m$; see the footnote on p. 93. Therefore when the position and slope of the line have been determined, the best points may at once be found graphically. If the length of the line joining the rth observed and best points is l_r, and the length of the perpendicular from the rth observed point to the line is d_r, then it is easily seen that $(x_r-X_r)^2=l_r^2\cos^2\theta=d_r^2/4\sin^2\theta$, and $(y_r-Y_r)^2=d_r^2/4\cos^2\theta$. Circles with centres (X_r, Y_r) and radii d_r are now drawn, and a planimeter is started at some point on the line to the left of the left-hand circle and made to trace the line and to pass clockwise round each circle as it comes to it, returning finally straight along the line to the starting point. The planimeter reading will give the total area of the circles, i.e.

$$\pi\Sigma d_r^2=\pi4\sin^2\theta\Sigma\,(x_r-X_r)^2=\pi4\cos^2\theta\Sigma\,(y_r-Y_r)^2.$$

When the number of observed points is large, the amount of computation saved by this method is considerable.

(*a*) *Error in centroid.* (i) *Risk for a distance* 0·01. The probability p, that the modulus of the error in $\bar{X}$ is not greater than $0{\cdot}01/\sqrt{2}$, is

$$\frac{2}{\sqrt{\pi}}\int_0^{0\cdot01/\surd 2} e^{-4h^2\mu^2}\,d\,(2h\mu)$$

from equations (23) and (24).

Writing $t=2h\mu$, we have

$$p=\frac{2}{\sqrt{\pi}}\int_0^{0\cdot01\times\surd2\times h} e^{-t^2}\,dt$$

$$=\frac{2}{\sqrt{\pi}}\int_0^{0\cdot97} e^{-t^2}\,dt=0{\cdot}83.$$

Similarly
$$p'=\frac{2}{\sqrt{\pi}}\int_0^{0\cdot725} e^{-t^2}\,dt=0{\cdot}69.$$

The probability that the true position of the centroid shall not be farther than 0·01 from $(\bar{X}, \bar{Y})$ is $pp'=0{\cdot}83\times0{\cdot}69=0{\cdot}57$.

I therefore take a risk of 1 in $1/(1-0{\cdot}57)$, i.e. 1 in 2·3, of being wrong by asserting that the true position of the centroid is not farther than 0·01 from $(\bar{X}, \bar{Y})$.

(ii) *Distance for a risk* 1 *in* 100.

$$1/(1-pp')=100, \quad pp'=99/100.$$

Now
$$p=\frac{2}{\sqrt{\pi}}\int_0^{2h\mu/\surd2} e^{-t^2}\,dt=I\,(2h\mu/\sqrt{2}), \quad \text{say}$$

and
$$p'=\frac{2}{\sqrt{\pi}}\int_0^{2h'\mu/\surd2} e^{-t^2}\,dt=I\,(2h'\mu/\sqrt{2}), \quad \text{say.}$$

Therefore
$$pp'=I\,(97\mu)\times I\,(72{\cdot}5\,\mu)=0{\cdot}99.$$

Hence μ is about 0·025 (shown in figure 1 as a dotted circle). I therefore take a risk of 1 in 100 of being wrong by asserting that the true position of the centroid is not farther than 0·025 from $(\bar{X}, \bar{Y})$.

(*b*) *Error in inclination.* From equations (39) and (40) the probability that the modulus of the error in m is not greater than ϵ is

$$=q'=2\int_0^{\epsilon}\chi'(\mu)\,d\mu, \quad \text{where} \quad \chi(\mu)=\frac{1}{\sqrt{\pi}}\,\frac{e^{-\mu^2/E'^2}}{E'},$$

and

$$E'^2=m^2\left(\frac{1}{h^2}+\frac{1}{h'^2}\right)=1{\cdot}813\,\frac{42+76}{2}\times10^{-5}=1{\cdot}07\times10^{-3},$$

$$E'=3{\cdot}27\times10^{-2}.$$

(1) *Risk for an error* 0·01. The probability that the modulus of the error in m is not greater than 0·01 is therefore

$$=\frac{2}{\sqrt{\pi}}\int_0^{\mu=0\cdot01} e^{-(0\cdot01)^2/E'^2}\,\frac{d\mu}{E'}$$

$$=\frac{2}{\sqrt{\pi}}\int_0^{0\cdot01/E'} e^{-t^2}\,dt=\frac{2}{\sqrt{\pi}}\int_0^{0\cdot316} e^{-t^2}\,dt$$

$$=0{\cdot}345.$$

I therefore take a risk of 1 in $1/(1-0{\cdot}345)$, i.e. 1 in 1·5, of being wrong by asserting that the value of m differs from 1·347 by less than 0·01. In this particular case an error of $\pm 0{\cdot}01$ in m corresponds to an error of about $\pm 13'$ in θ.

(2) *Error for a risk* 1 *in* 100. If we make $q'=0{\cdot}99$, we have the equation

$$\frac{2}{\sqrt{\pi}}\int_0^{\mu/F'} e^{-t^2}\,dt=0{\cdot}99 \text{ to determine } \mu.$$

Hence $\mu/F'=1{\cdot}82$, therefore $\mu=1{\cdot}82\times 3{\cdot}27\times 10^{-2}=0{\cdot}06$.

Therefore I take a risk of 1 in 100 of being wrong by asserting that the error in m is less than 0·06. In this particular case an error of $\pm 0{\cdot}06$ in m corresponds to an error of about $\pm 1^\circ\ 15'$ in θ (shown in figure 1 by dotted lines).

§ 10. THE ASSUMPTION OF GAUSSIAN ERROR LAWS

The question now arises as to whether repeated observations of physical quantities actually do obey error laws of the Gaussian type. This is a point which can only be decided by experiment in each particular case. Three different cases present themselves.

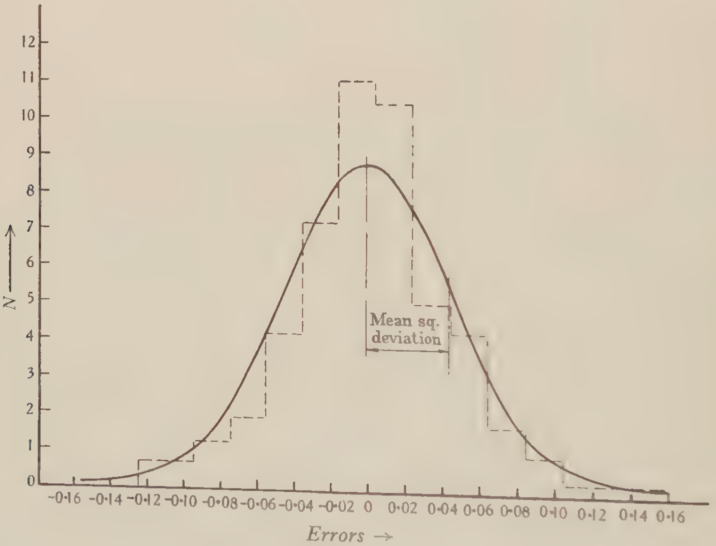


Figure 3. Copper loss. Mean, 2·62 %; h, 15·6; Mean-square deviation, $0{\cdot}04_5$.

(i) If repeated observations of the two coordinates of any one point can be made separately, then the results will show whether Gaussian error laws are obeyed and, if so, what are the values of the precision constants. These values will suffice for the solution of the problem dealt with in this paper, provided it can at the same time be assumed that the error laws for all the X's are the same and that the error laws for all the Y's are the same.

(ii) If repeated observations of any one point can be made, but it is not possible to keep one coordinate constant while making repeated observations on the other, it will still be theoretically possible to make a three-dimensional model of the frequency distribution. By taking sections along the X and Y axes in turn, the types of error law may be ascertained and their parameters determined.

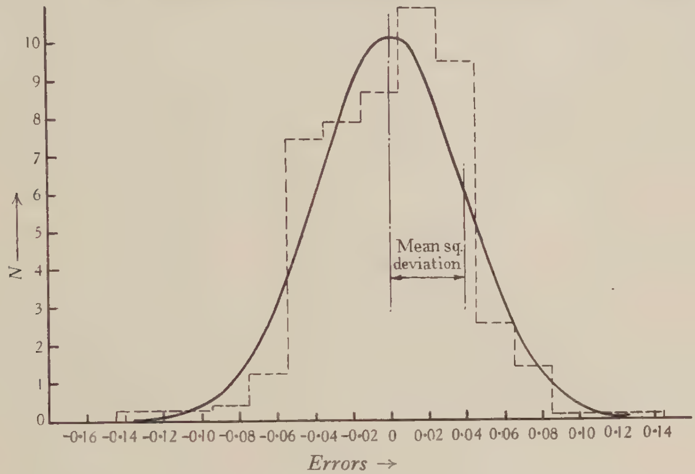


Figure 4. Impedance voltage. Mean, 3·14 %; h, 18·1; Mean-square deviation, 0·039.

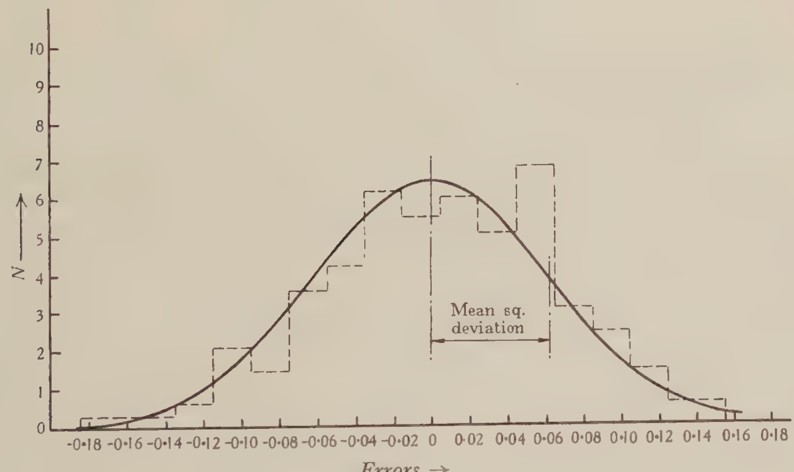


Figure 5. Iron loss. Mean, 2·11 %; h, 11·5; Mean-square deviation, 0·062.

(iii) If it is impossible to make any repeated observations at all, the only possible test is to examine the errors $(x_r' - X_r')$ and $(y_r' - Y_r')$ as soon as x_r' and y_r' have been determined. If a sufficiently large number of points have been observed, these errors should be distributed according to the Gaussian law, using the most probable values of the precision constants. If it is found that they are not so distributed, the method of the foregoing pages does not apply.

It may be noted here that this test should be made as a check even when repeated observations are possible.

Three examples of frequency-distribution curves of errors which do actually obey the Gaussian law are given in figures 3–5. They were obtained from the measurement of the percentage copper loss, impedance volts and iron loss of 310 transformers of identical design and made in the same workshop. In each case the ordinate N is the number of occurrences divided by the product of the total number of observations and the interval, the interval being 0·02 per cent. The broken lines represent the observed deviations from the mean, and the theoretical curves are drawn so as to have the same mean-square deviations as the observations. It is seen that the fit is fairly good in figure 3 and quite good in figures 4 and 5.

§ 11. ACKNOWLEDGMENT

I have great pleasure in thanking Mr A. P. M. Fleming, C.B.E., Director of the Metropolitan-Vickers Electrical Co. Ltd. and Manager of the Research Department, for permission to publish this work.

DISCUSSION

Dr W. Edwards Deming (*communicated*). This paper brings up a host of interesting details. In the first place, bits of history concerning the subject are continually coming to light. It seems that R. J. Adcock in 1878* first worked on the problem of fitting a line to a set of points by minimizing the sum of the squares of the normals dropped upon the line. That he made a curious slip in his algebra was pointed out by Kummell (*infra*). The general problem of curve-fitting by the method of least squares was outlined by Charles H. Kummell in 1879†. He obtained the correct solution for the straight line under the condition that the x and y coordinates have constant but not necessarily equal weights, and he showed how to determine the weights of the adjusted parameters in both this and the non-linear problem. Mansfield Merriman in 1890‡ also worked on the straight line, but made no new contribution to theory. The next advance was made by Karl Pearson in 1901 in his memoir "On lines and planes of closest fit to systems of points in space§," wherein both the best and worst-fitting lines and planes were obtained, together with ingenious and now well known formulae for the mean-square residual, the premise being that all coordinates have equal weight. Apparently unaware of earlier work, Corrado Gini in 1921‖ found the best and worst-fitting lines for points whose x and y coordinates have constant but not necessarily equal weights. Gini's

* *Analyst* (Des Moines), **5**, 53–54 (1878). † *Analyst*, **6**, 97–105 (1879).
‡ *Report of the United States Coast and Geodetic Survey*, p. 687 (1890).
§ *Phil. Mag.* **2**, 559–572 (1901). ‖ *Metron*, **1**, No. 3, 63–82 (1921).

aper has, in turn, been overlooked by most subsequent writers, among whom I ave been numbered.

Most of the investigators who have attempted to take account of errors in both and y observed coordinates have worked on the straight line, and have recognized he impossibility of getting an exact solution without making the assumption that he x coordinates all have equal weight and the y coordinates all have equal weight. Unfortunately this assumption divests the problem of much that promised to be of nportance, for it is a fact that lines obtained with various values of the ratio h/h' in the author's notation) differ only in the squares of the residuals. Now since in ractice the residuals are usually small, it turns out that the lines thus obtained usually lo not differ significantly. Miss Dent's drawings afford a good illustration of this oint. The non-linear relation does not exhibit this peculiarity, nor does the inear relation, save under the assumption of constant x- and constant y-weighting.

These facts are closely related to the author's observation that Mr Cook's method of estimating h/h' is applicable only to equations of the second or higher degree. Mr Cook's method breaks down in the case of the straight line only in the sense that t requires to be carried to a second approximation, which is just what one would expect in view of the fact that the ratio h/h' becomes of secondary importance when he precisions of the x and y coordinates are each constant.

It is interesting to note that equations (33) and (34) obtained by the author for estimates of h and h' differ from the usual ones in having s where one usually sees $(s-2)$. The author's formulae were based on the assumption that the horizontal and vertical line segments between the observed and calculated points are actual errors, whereas they are usually considered as residuals. The "number of degrees of freedom," as it was named by R. A. Fisher, is s if these line segments are errors, and $(s-2)$ if they are residuals.

Miss Dent recognizes the importance of testing the calculated inclination and intercept of the fitted line. From the work of Helmert, "Student," Fisher, and others, it is known that if the errors of observation are normally distributed the values of the calculated precision constants will not be so; hence the equations of §6 for testing the precision of the slope and intercept of the fitted line must be regarded as approximations which, however, improve as the number of observed points increases. Henry Schultz* has worked out a complete system for estimating the standard errors of a curve and its parameters. In this connection, Chapter 5 of R. A. Fisher's *Statistical Methods for Research Workers* should not be overlooked. The risk of 1 in $1/(1-pp')$ for an error μ in the centroid, as given by the author in §6, is more restrictive than necessary, but a full treatment would necessitate opening for discussion the whole subject of statistical tests.

Finally, a word should be said regarding the normal error law†. Even when it is not obeyed—and it probably never was—statistical deductions based on sampling from a normal parent population are not apt to lead one into wrong conclusions,

* *J. Amer. stat. Ass.* 25, 139–185 (1930).

† Commonly called Gaussian; see, however, footnote 8 in a paper by W. Edwards Deming and Raymond T. Birge in *Reviews of Modern Physics*, 6, 119–161 (1934).

provided reasonable care is exercised. But if one is interested in testing the normal law, or any other law, the approach might well be made with some objective criterion such as Karl Pearson's chi test.

Author's reply. I should like to thank Dr Deming for his interesting historical notes.

In comparison with any method which requires to be carried to a second approximation, or which involves the preliminary determination of approximate values for the parameters, the method of the present paper is much simpler and shorter.

In § 6 the values of h and h' are regarded as accurately known. It seems doubtful whether further approximations, involving the errors in the calculated precision constants, would be justified.

537·533·706 548·74

OXIDE FILMS ON LIQUID METALS STUDIED BY MEANS OF ELECTRON-DIFFRACTION

By R. O. JENKINS, A.R.C.S., B.Sc.,
Imperial College of Science and Technology

Received July 16, 1934. *Read November* 2, 1934.

BSTRACT. Electron-diffraction patterns obtained by reflection from molten lead, nc, bismuth and tin have shown that the surfaces of these metals are covered with thin xide films. From these patterns it has been deduced that the films consist of small ectron-optically flat oxide crystals resting on their [001] faces. The inner potentials and ystal structures of these oxides have been found, and they have thus been identified as efinite chemical compounds. It has also been found possible to remove these films from e molten metal by means of nickel gauze and to obtain electron-diffraction patterns by ansmission. The transmission patterns indicate that although the removed films are f the same chemical composition as those on the molten metal, their removal has disurbed the orientation and given the crystals an approximately Maxwellian distribution f orientations about their original direction.

§ 1. INTRODUCTORY

THE work to be described in this paper was the outcome of an attempt to study the surface of liquid metals by means of electron-diffraction. The metals used, which had to be of fairly low melting-point, all acquired quite thick oxide layers when melted in air, and became oxidized to some extent even in a fairly good vacuum. During an attempt to eliminate these oxide films, curious orientations of the crystals comprising the films were observed by means of electron-diffraction. The experiments to be described consisted mainly in photographing electron-diffraction patterns by reflection from the oxide-covered surface of the molten metal, and identifying the oxide and its mode of orientation.

In addition to this, experiments were carried out by means of a new method of removing the oxide film from the surface of the metal so that that electron beam might be diffracted by transmission through the thin film. These oxide films were formed on the metal in air, while those used for reflection were formed *in vacuo*. The structure and orientation however seem to be similar in the case of the metals tried, except that the orientation was not complete in the transmission specimens.

§ 2. EXPERIMENTAL APPARATUS

The diffraction camera used for this work was one of the Thomson-Fraser type[1] with modifications designed for Prof. Thomson by C. G. Fraser of Aberdeen, and set in a horizontal plane to obtain the electron beam at grazing incidence on the

liquid surface. The discharge tube had the usual aluminium disc cathode, but its other end was coned and ground to fit a corresponding ground cone on the apparatus. The joint was made vacuum tight with Apiezon M grease and cooled by a water jacket round the outside, figure 1. The electron beam was defined by the anode which consisted of a hollow aluminium cylinder with an aluminium plate round the outside at the discharge-tube end. Each end of the hollow cylinder was closed by a molybdenum disc pierced with a hole 0·15 mm. in diameter, which allowed a fine electron beam to pass from the discharge tube to the camera. Directly in front of the anode was an aluminium shutter operated from the outside of the specimen chamber through a coned, ground and greased spindle, figure 2, which also operated an interconnected razor-blade scraper just above the surface of the specimen, figure 1.

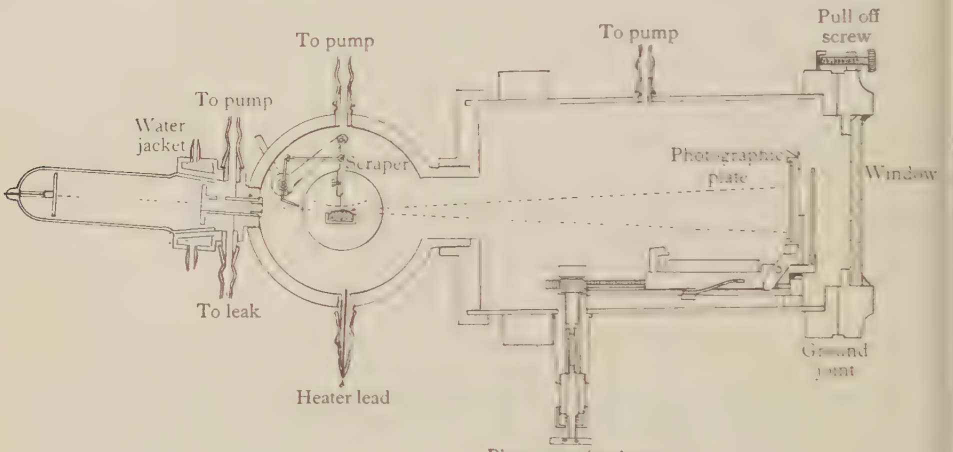


Figure 1. Diagram of apparatus.

The method of carrying the specimen may be seen from figure 2. By turning the large wheel on the outside of the specimen-chamber the specimen could be rotated in azimuth by a gear comprising two large bevel wheels, while by operating the smaller lever, a rack and pinion raised or lowered the specimen across the electron beam. Each of these movements worked independently through the two small coned and ground spindles. In addition, the angle that the specimen surface made with the beam could be altered by turning the large coned and ground joint which carried the whole specimen assembly.

The specimen was removed from the apparatus by breaking the flat ground and greased joint round the outside edge of the cylindrical specimen-chamber. This joint being nearly 2 cm. wide and about 15 cm. in diameter, considerable force was needed to break it. This was applied by means of a specially designed screw clamp after the safety ring had been removed. When the specimen was in position inside the apparatus it could be observed through a plate-glass window sealed with picein vacuum wax to the far side of the specimen-chamber.

Access to the photographic plate and its carrier was by means of the end cover the camera, which also had a large flat ground and greased joint with drawing-off rews, figure 1. To this cover was waxed a plate-glass window through which the ecimen and fluorescent screen could be observed. The plate-holder consisted of carriage running on two rails fixed to the body of the camera and it could be moved ackwards and forwards by means of a rack and pinion worked through a conical eased and ground spindle. The photographic plate was carried by the lid of a

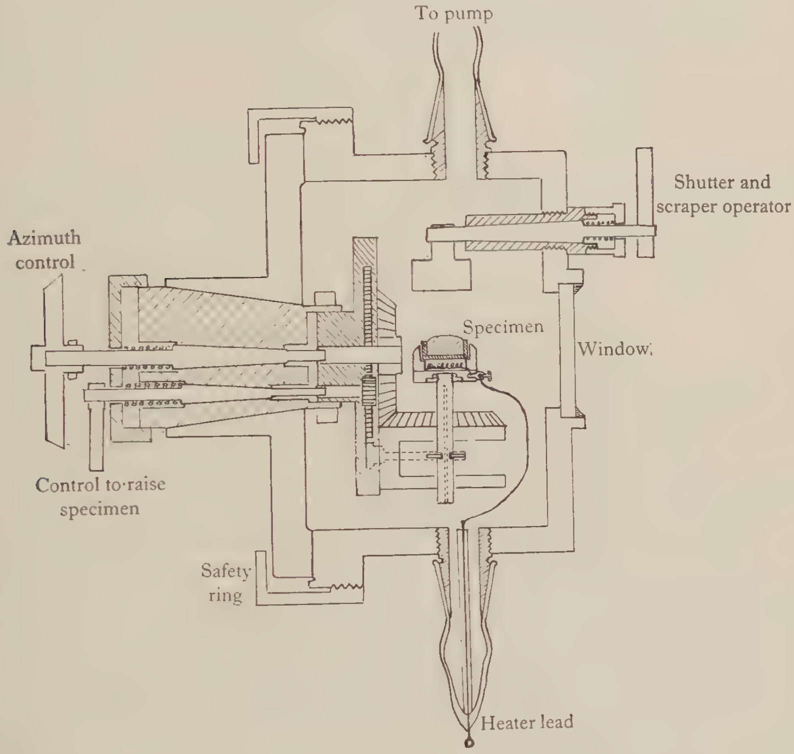


Figure 2. Section through specimen-chamber excluding razor-blade scraper.

light-tight cassette, usually held down by a spring catch in a horizontal position on o the rest of the cassette so as to allow the diffracted beams to hit the fluorescent screen. When the carriage was racked towards the end of the camera, a rod in contact with the catch came up against a stop at the end of the rails and released the catch, allowing the cassette lid carrying the plate to spring up into the vertical position as shown in figure 1. After the pattern had been photographed the plate carriage could be racked back along the rails, when the two projecting lugs on the cassette lid each came in contact with a strong spring blade and closed the lid down again to the horizontal position.

§ 3. VACUUM AND HIGH TENSION EQUIPMENT

The whole apparatus was evacuated through wide-bore glass tubing connected through a liquid-air trap to a mercury diffusion pump backed by a Cenco-Hyvac pump. The connections to the apparatus were made by means of standard taper coned ground and greased joints, one for each of the three sections of the apparatus, figure 1. Under working conditions the pressure in the discharge tube should be about 10^{-3} mm. of mercury and that in the camera rather lower. The difference in pressure was obtained by allowing air or hydrogen to leak into the discharge-tube through an adjustable needle valve from a bottle reservoir maintained at a pressure of about 4 mm. of mercury.

The high-tension current for the discharge was obtained from an induction coil with mercury interrupter. The current was passed through a rectifying valve and water resistance to the discharge, while a smoothing-condenser of capacity $0{\cdot}004\,\mu$F. was connected from the anode of the valve to earth. The discharge voltage was measured by spark-gap between 5 cm. spheres. By running the interrupter at high speed it was found possible to obtain a very homogeneous electron beam with this equipment.

§ 4. SPECIMEN-HEATER

The specimen-heater used in conjunction with the camera consisted of a circular asbestos cement base on a brass plate which screwed on to the specimen-pillar, figure 2. The heater coil of nichrome wire carrying about 2 amperes at 30 volts rested on the asbestos base and had one end connected to the brass plate and the other to a terminal. The crucible containing the molten metal rested in a ring of asbestos cement fixed to the base by two screws. After preliminary tests, crucibles of Acheson graphite turned on a lathe were used, as they would not charge up when the electron beam hit the metal, and the graphite would not contaminate the metal. The terminal on the outside of the heater was connected by a flexible insulated lead to a wire sealed through the end of a glass tube connected to the bottom of the crystal chamber, figure 2.

§ 5. THEORY

The electron beam has a wave-length associated with it (De Broglie) and can therefore be diffracted by crystals in a similar manner to X-rays according to
θ, d Bragg's law $n\lambda = 2d\sin\theta$, where θ is the angle of diffraction and d the lattice spacing of the crystal planes. Combining these two equations and the camera-dimensions, we find that

$$d = \frac{L\sqrt{150}}{r\sqrt{v}} \times 10^{-8}\ \text{cm.} \qquad \ldots\ldots(1),$$

L, v, r where L is the camera-length, v the voltage of the discharge and r the distance from the diffracted spot to the undeviated spot on the photographic plate.

If the surface is electron-optically flat, so that electrons enter and leave the specimen in planes parallel to the surface, refraction takes place as a result of the difference in potential between the inside and outside of the specimen. This is

equivalent to a refractive index μ equal to $\sqrt{(1+\phi/v)}$[2], where ϕ is the inner potential. This alters all component distances of diffracted spots in a direction perpendicular to the shadow-edge. It may be shown that if the angles made with the specimen-surface by the incident and diffracted beams are equal μ, ϕ

$$y^2 = y_0^2 - 4L^2\phi/v \qquad \text{......(2)},$$

where y is the observed component distance of r along a line perpendicular to the specimen-surface, and y_0 is what the distance would be if no refraction had taken place. If a rotation picture is taken and a number of orders of the same plane appear on the line perpendicular to the specimen-surface through the undeviated spot, equation (2) may be put into the form y y_0

$$y^2 = n^2 r_0^2 - 4L^2\phi/v \qquad \text{......(3)},$$

where n is the order number and r_0 the distance from the first-order spot to the undeviated spot, as calculated from the Bragg angle. If y^2 is plotted against n^2, the slope of the graph will give a value for r_0 which may be substituted in equation (1) to give the lattice spacing d, while the intercept will give a value of the inner potential ϕ. n, r_0

The patterns obtained by reflection from the oxide films may be explained as follows. If a single crystal is placed in the beam with a cleavage face as surface, the resulting pattern on the plate consists of spots at the intersection of three diffraction curves as follows. The atoms on a line in the surface perpendicular to the electron beam give a series of equidistant straight lines perpendicular to the surface of the specimen. Even if the crystal is very small these lines will be quite sharp. The atoms lying in the surface along the beam give a series of circles. Owing to the small angle this line makes with the beam, these circles may become quite broad if the crystal is very small. Finally the planes parallel to the surface give a series of lines parallel to the surface, spaced according to Bragg's law as modified by the refraction effect. There will be a spot wherever these three diffraction curves intersect. If, however, the crystal surface may be considered flat and continuous for only a small area the second curve becomes very diffuse, and if the scattering atoms are heavy the penetration of the electrons will be small and the third condition will be relaxed to some extent. These effects have been discussed by Kirchner and Raether in some detail[3]. In the patterns under consideration the crystals are small enough to make the second diffraction condition so relaxed that its effect may be ignored, and the pattern thus consists of spots at the intersection of the lines perpendicular and parallel to the shadow-edge on the photographic plate. Owing to the relaxation of the third condition these Bragg spots are elongated along the line perpendicular to the shadow-edge, so that the lines are continuous but are of greater intensity where a spot should occur. The meniscus of the surface gives the equivalent of a rotation picture and a large number of spots may thus appear, while the side lines correspondingly may be quite long. If in addition to alteration of the angle of incidence the crystal is rotated in azimuth about an axis perpendicular to its surface, the atomic distances along the line perpendicular to the electron beam will change, and a series of side lines corresponding to all the possible atomic distances in the surface will be obtained.

If the [001] plane is in the surface these distances are the lattice spacings of the [*xy*0] planes, where x and y may have any integral values. The series of spots on one particular side line due to spacing [*hj*0] will correspond to the zone of planes [*hjz*], where z will have any integral value.

With this theory as a basis it has been possible to evaluate two of the axes of the oxide crystals from the distances apart of the diffraction side lines. The third axis and inner potential have been obtained by applying equation (3) to the measurements on the line through the undeviated spot. From the length of the axes the oxides have been in most cases identified by comparison with X-ray crystal structures. It has thus been shown that the oxide surfaces consist of quite small crystals usually resting on their [001] faces on the liquid metal surface.

§ 6. EXPERIMENTAL RESULTS

Lead. Pure lead was melted in air in a graphite crucible and the surface was skimmed. After cooling it was placed in the apparatus, which was evacuated. The current in the heating-coil was then switched on, so as to melt the metal and maintain it at a temperature not very far above its melting-point. Owing to degassing of the heater it was necessary to wait for a few minutes till the vacuum had recovered sufficiently to allow a diffraction pattern to be taken. The specimen was then raised completely across the electron beam till the swinging razor-blade scraper could traverse the surface when the operating lever was moved. In this way a clean patch of metal was obtained. The specimen was lowered again till the electron beam could graze the surface, when the diffraction pattern could be observed on the fluorescent screen and subsequently photographed.

The complexity of the pattern suggested that the oxide had not one of the simplest forms of symmetry, figure 3. A complete analysis of the pattern was made and it was deduced that the oxide was the rhombic form of PbO known as litharge, which has axes a, b, c equal to 5·50 A., 4·68 A., and 5·88 A. respectively, all the small crystals being arranged with their [001] planes parallel to the surface of the specimen. The definition of the spots on the lines was bad owing to the smallness of the crystal-size and lack of penetration of the electron beam, lead being a heavy atom. The measurement of the spots on the central line was thus not very accurate, but the value of the c axis of the crystals calculated from these measurements agreed within 1½ per cent with the X-ray value.

As the a and b axes of the crystals were unequal, the side lines corresponding to the [*jhz*] planes would be different from those of the [*hjz*] planes, and this would make the number of lines very large. Some of the lines did not however appear owing to the form of the space group in the unit cell. From the X-ray study of litharge by the Debye-Scherrer ring method(4), the absence of the rings due to the [10*z*] [01*z*] [21*z*] [12*z*] [12*z*] [30*z*] [03*z*] [32*z*] and [23*z*] planes meant that the corresponding side lines would be missing, so that the pattern would be simplified. Table 1 gives a summary of measurements made on plates obtained with lead oxide.

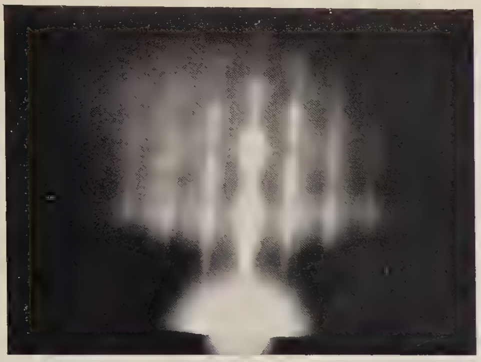


Figure 3. Reflection pattern from litharge on molten lead.

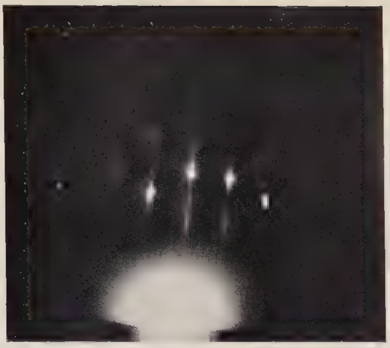


Figure 4. Reflection pattern from litharge on lead crystal.

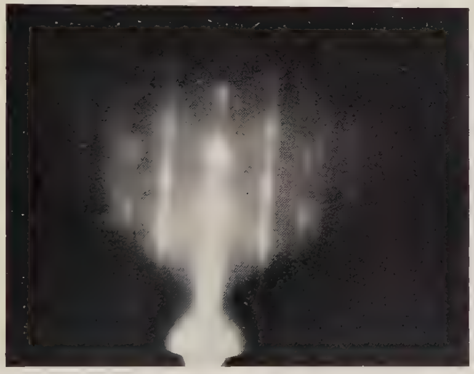


Figure 5. Reflection pattern from zinc oxide on molten zinc.

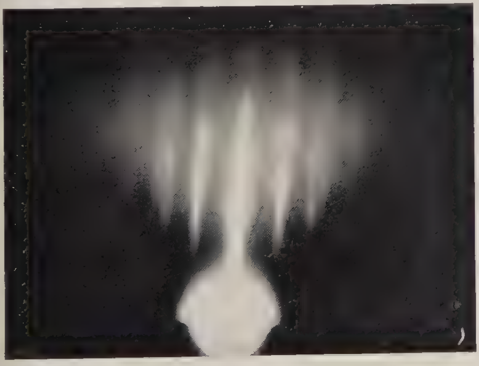


Figure 6. Reflection pattern from bismuth oxide on molten bismuth.

Figure 7. Reflection pattern from tin monoxide on molten tin.

[To face p. 114

In addition to the side lines tabulated, a very faint side line inside the [11z] line seemed to correspond to the forbidden [10z] series of planes.

Table 1

Indices of zone of planes causing side line	Calculated lattice spacing, $z=0$ (A.)	Observed lattice spacing, $z=0$ (A.)	Intensity
11z	3·57	3·58	Very strong
20z	2·75	2·78	Medium
02z	2·34	2·36	Medium
22z	1·79	1·79	Strong
31z	1·71	1·71	Medium
13z	1·51	1·52	Medium
40z	1·37	1·37	Very weak
33z	1·19	1·20	Medium

The mean value of the *c* axis was 5·97 A. and the inner potential 12·3 V.

The value for the *c* axis is about $1\frac{1}{2}$ per cent too high, but some of this error may be due to the thermal expansion of the crystals, which would slightly increase all the lattice spacings. Even if this effect is neglected, the value is within the limit of experimental error. The figure obtained for the inner potential is probably only accurate to about one volt.

Assuming that the small crystals comprising the layer had rotational freedom round the *c* axis when the metal was liquid, it was of some interest to see whether this was still the case when the metal underneath was of crystalline form, or whether any orientation took place on crystallization of the metal. This question was tested by allowing the lead to cool down very slowly by decreasing the current in the heater coil, when the lead would crystallize at the surface into a few large flat crystals from whose surfaces diffraction patterns could be obtained. It was found that on solidification the small oxide crystals on the lead had all tended to set in one direction on the top of the metal crystal, for true single-crystal diffraction pictures due to the oxide were obtained. These showed plainly only one set of equally spaced side lines with sharp spots on them, placed according to the other two diffraction conditions, figure 4. It can be seen that the second diffraction condition was obeyed rigidly, all the spots appearing on a circle. This shows that the surface must have been flat and continuous for a much larger area than in the case of the small crystals on the liquid metal. In addition to these spots there were Kikuchi lines which would also appear only with a true single crystal. It thus appears that on crystallization of the underlying metal the small oxide crystals turn so that they all are in the same azimuth, and thus make themselves into a layer of true single-crystal form on the top of the metal crystal.

Attempts were also made to reduce the oxide layer by means of atomic hydrogen, so as to leave a clean metal surface. A discharge tube of the Woods type was waxed in position through a hole in the inspection window on the far side of the specimen-chamber, and was arranged to run at a pressure of about $\frac{1}{10}$ mm. of mercury, being

supplied from a bottle reservoir of hydrogen at a pressure of a few centimetres through an adjustable leak. The atomic hydrogen made by the discharge escaped through a pinhole close to the specimen-surface where it was possible that it might reduce the oxide. The oxide was not reduced but underwent rather interesting and striking changes. After the atomic-hydrogen tube had been run for about a minute and the diffraction pattern from the surface had been observed directly the discharge was cut off, two entirely new patterns were obtained. These both were due to crystals with a body-centred tetragonal lattice resting on their [001] faces, while these crystals were not electron-optically flat, no refraction effect being observed. The approximate values of the axes for these crystals were as follows: $a=4{\cdot}10$ A., $c=6{\cdot}00$ A. for one pattern, and $a=3{\cdot}85$ A., $c=13{\cdot}30$ A. for the other. Neither of these structures correspond to a known lead compound, while the second one suggests something fairly complex owing to its very large c axis.

Zinc. Pure zinc was melted in a graphite crucible and after cooling was placed in the apparatus. The procedure for melting and skimming the surface inside the apparatus was the same as that used for lead. It was found that molten zinc had a considerable tendency to vaporize, necessitating cleaning of the window and metal parts in the specimen-chamber after the taking of a diffraction photograph, and for this reason only a few plates were taken. These however gave quite definite information as to the chemical nature and orientation of the oxide surface. The film consisted of zinc-oxide (ZnO) crystals of ordinary close-packed hexagonal structure ($a=3{\cdot}22$ A., $c=5{\cdot}18$ A.) resting on their [001] faces, with their c axes perpendicular to the surface.

In the case of hexagonal crystals oriented in this manner the side lines due to the [hjz] planes would correspond to lattice spacings $a\sqrt{3}/2(h^2+hj+j^2)^{\frac{1}{2}}$, while owing to the fact that the crystals were of close-packed structure many of the Bragg spots would not appear on the side lines. The spots in the case of zinc were rather better defined than those from lead, because the lighter zinc atoms allowed greater electron penetration and resolving power, figure 5.

The measurements for zinc oxide are given in table 2 and show quite good agreement with X-ray values, besides giving a figure for the inner potential of the oxide correct to approximately one volt.

Table 2

Indices of zone of planes causing side line	Calculated lattice spacing, $z=0$ (A.)	Observed lattice spacing, $z=0$ (A.)
$10z$	2·79	2·79
$11z$	1·61	1·62
$20z$	1·39	1·40
$21z$	1·05	1·06
$30z$	0·93	0·93
$22z$	0·80	0·81

From measurements on the central line $c=5{\cdot}27$ Å. [X-ray value 5·18]. Inner potential $\phi=12{\cdot}3$ V.

The lattice spacings from the side lines and from the centre line are again slightly high, but they are just within the limits of experimental error even when thermal expansion is neglected. In addition to the side lines tabulated, a very faint one was observed inside the [10z] line which gave a lattice spacing for $z=0$ of approximately 3·86 A. This does not appear to bear any simple relation to the other side lines and may be caused by an oriented impurity. It may be connected with an extra ring obtained by Finch and Quarrell(5) who used electron-diffraction by transmission through an incompletely oriented film of zinc oxide. This ring gave an observed lattice spacing of 3·79 A. and was tabulated as a half order of the [102] plane of zinc oxide. It might however be the ring due to the first plane of the zone causing the side line by reflection, while the rings due to the rest of the zone of planes would be of greater radius and less intensity and might not be strong enough to appear on the plate.

Bismuth. The oxide surface on bismuth was prepared in exactly the same way as for the previous metals, the metal being used at a temperature not far above its melting-point. The resultant pattern, figure 6, showed that the oxide had a hexagonal structure with the bismuth atoms placed in the unit cell in approximately the same positions as the zinc atoms in the unit cell of zinc oxide, and that the crystals of the oxide layer were all resting on their [001] faces. The axes have been calculated from the measurements but the oxide has not been identified chemically because no crystal structures of bismuth oxides are given in Landolt and Börnstein's tables or Ewald and Herrmann's *Structurbericht*(6). As bismuth is trivalent the oxide would probably be of the formula Bi_2O_3. The other oxides of this trivalent type having a hexagonal structure are those of the rare-earth group such as cerium. The structure of compounds of this type [D 52] places the metal atoms in a close-packed hexagonal lattice with an axial ratio of 1·56 and an a axis of approximately 3·90 A. These figures agree very closely with those obtained for the bismuth oxide which are 1·57 for the axial ratio and 4·02 A. for the a axis. The suggested structure is made even more probable by the fact that the ionic diameters of the rare earths and bismuth are nearly the same.

In addition to the side lines given in table 3, two very faint lines were observed

Table 3

Indices of zone of planes causing side line	Mean observed lattice spacing $z=0$ (A.)	Corresponding value of a axis (A.)
10z	3·47	4·01
11z	2·02	4·04
20z	1·73	4·00
21z	1·32	4·03
30z	1·16	4·02
22z	1·01	4·04
31z	0·965	4·02

Mean value of $a=4·02$ A. From central line $c=6·30$ A. Hence $c/a=1·57$.
Mean value of inner potential $\phi=11·0$ V.

inside the [10z] line which did not appear to bear any simple relation with the main pattern. These again might be due to an oriented impurity as suggested in the case of zinc.

Tin. An oxide surface on pure tin was prepared as with the previous metals, the tin being used at a temperature just above its melting-point. The resultant pattern, figure 7, when analysed showed that the film consisted of the tetragonal form of SnO [$a=3{\cdot}80$ A., $c=4{\cdot}81$ A.], all the crystals resting on their [001] faces. *u* The tin atoms are placed in the unit cell in positions (000) and ($\frac{1}{2}\,\frac{1}{2}\,2u$), where u is an undetermined parameter. As all orders of the [001] plane appeared on the central line u cannot be $\frac{1}{4}$, which would place the tin atoms in a body-centred lattice. The side lines obtained from the tetragonal lattice corresponded to planes [hjz], where the lattice spacing for $z=0$ was $a/(h^2+j^2)^{\frac{1}{2}}$. The mean measurements for these side lines are given in table 4 and show quite good agreement with the calculated values. The value for the c axis is about 2 per cent too high, which is just within experimental error due to the diffuseness of the spots caused by lack of penetration of the electrons. The inner potential obtained from the measurements would also be correct only to about one volt.

Table 4

Indices of zone of planes causing side line	Strength of side line	Observed lattice spacing, $z=0$ (A.)	Calculated lattice spacing, $z=0$ (A.)
10z	Strong	3·80	3·80
11z	Strong	2·68	2·69
20z	Medium weak	1·90	1·90
21z	Medium	1·70	1·70
22z	Medium weak	1·35	1·34
30z	Weak	1·28	1·27
31z	Strong	1·21	1·20
32z	Strong	1·05	1·05

From measurements on central line $c=4{\cdot}91$ A.
Mean inner potential $\phi=14{\cdot}2$ V.

In addition to that already explained, an entirely different pattern also was obtained from the tin surface at a slightly higher temperature. This pattern, while apparently due to an oriented layer of character rather similar to that of the layer of SnO obtained previously, was considerably more complex and while some of its main lattice spacings have been evaluated it has not been connected with any definite oxide of tin with a particular orientation.

§7. OBSERVATIONS OF DIFFRACTION BY TRANSMISSION

These results having been obtained by means of the reflection arrangement, it was thought desirable to observe diffraction patterns by transmission through thin films of oxide. This has already been done for some metals with films prepared in different ways. Ponte(7) obtained films of zinc oxide showing no orientation, by burning zinc in air. Bragg and Darbyshire(8) obtained films by drawing a loop of

wire out of the molten metal, thus leaving a film of metal across the loop. In this film were holes across which thin membranes of oxide suitable for transmission diffraction were stretched. Using this method they obtained films of SnO_2, PbO_2 and ZnO. Films of zinc oxide prepared in this way have been studied in more detail by Finch and Quarrell (5), who showed conclusively that although the diffraction patterns resembled the face-centred cubic type they were really due to the ordinary close-packed hexagonal structure of ZnO, with the intensities of the rings greatly altered by the partial selective orientation of the crystals in the film. As these facts applied to zinc oxide it seemed possible that a similar pattern might be obtained from a film of bismuth oxide, the two reflection patterns being very similar, and experiments were tried to see whether this was so.

Bismuth-oxide films. While the loop method of preparing a specimen may be easily carried out for zinc, which will wet other metals, bismuth presents a more difficult problem. Alternative methods were therefore tried, one of which has proved very satisfactory both for bismuth and for other metals. A piece of fine nickel gauze having about 25 meshes per centimetre was dipped under the freshly skimmed surface of the metal and then lifted out with a pair of tweezers, the gauze being kept nearly horizontal. This procedure allowed the metal to drain away through the gauze, leaving patches of oxide film caught in the meshes. The specimen so prepared was then fitted in a clamp in the apparatus in place of the furnace used for the previous part of the work.

The resultant diffraction pattern was in fact almost precisely similar in appearance to that obtained by Finch and Quarrell for zinc oxide. The lattice spacings for the rings were worked out and the indices allotted with the help of Hull and Davey's nomograms (9) which were also used in the same way for the tin oxide diffraction patterns. The rings fitted a hexagonal close-packed lattice of axial ratio 1·56 and an *a* axis of 3·93 A. in a condition of partial orientation so that the greatest number of crystals had their [001] planes in the plane of the specimen and none at right angles to this direction. The mean lattice spacings are given in table 5, together with

Table 5

Indices of ring	Intensity of ring	Observed lattice spacing (A.)	Calculated lattice spacing (A.)
110	Very strong	3·39	3·40
101	Strong	2·92	2·97
102	Weak	2·26	2·27
110	Very strong	2·01	1·97
103	Very weak	1·77	1·75
200	Strong	1·71	1·70
201	Medium	1·64	1·64
120	Medium	1·30	1·29
121	Medium	1·26	1·26
300	Medium	1·16	1·13
302	Weak	1·08	1·06
220	Very weak	1·00	0·99
304	Medium	0·94	0·91

the indices and intensities and calculated lattice spacings assuming the values for the lattice which gave the best fit on the nomogram.

The axial ratio from the transmission observations agrees quite closely with that obtained by reflection (transmission 1·56, reflection 1·57) while the *a* axis is about 2 per cent lower by transmission than by reflection, and some of this difference may be accounted for by thermal expansion. The transmission value of 3·93 A. is almost exactly equal to that of other members of the same structure group and is therefore probably the more correct value at ordinary temperature.

Tin-oxide films. Thin films of tin oxide were also made by the nickel-gauze method. The tin was maintained at a temperature slightly above its melting-point and the surface was skimmed before the gauze was dipped. The resultant diffraction pattern showed that the film made in this way consisted almost entirely of the monoxide SnO, the crystals being partially oriented with their [001] faces in the plane of the film as in the case of bismuth and zinc. Here again, no crystals had their *c* axes lying in the plane of the film, for no orders of the [001] planes were observed. The greatest number of crystals had their *c* axes nearly perpendicular to the plane of the film, since rings with the third index zero were greatly strengthened, figure 8. Rather similar patterns have been obtained by Steinheil [10] from specimens made by playing a flame on tin foil. In this case the orientation was not so marked, as the [002] ring appeared with moderate intensity. The calculated and observed lattice spacings of the rings together with their intensities are given in table 6.

Table 6

Indices of ring	Intensity of ring	Observed lattice spacing (A.)	Calculated lattice spacing (A.)
100	Very weak	3·75	3·80
SnO_2 110	Very weak	3·39	3·34
101	Medium	3·02	2·98
110	Very strong	2·71	2·69
111	Weak	2·36	2·35
200	Strong	1·92	1·90
201	Medium strong	1·79	1·77
120	Weak	1·70	1·70
121	Strong	1·62	1·60
202	Weak	1·51	1·49
122	Weak	1·42	1·39
220	Medium strong	1·35	1·34
221	Weak	1·31	1·29
203, 301	Medium	1·24	1·22
130	Medium	1·20	1·20
123	Weak	1·18	1·17
302	Very weak	1·13	1·12
132	Medium	1·08	1·08
231, 223	Medium	1·04	1·03

In addition to films consisting of the monoxide, some composed of the dioxide SnO_2 were made in a similar way with the tin at a rather higher temperature. This

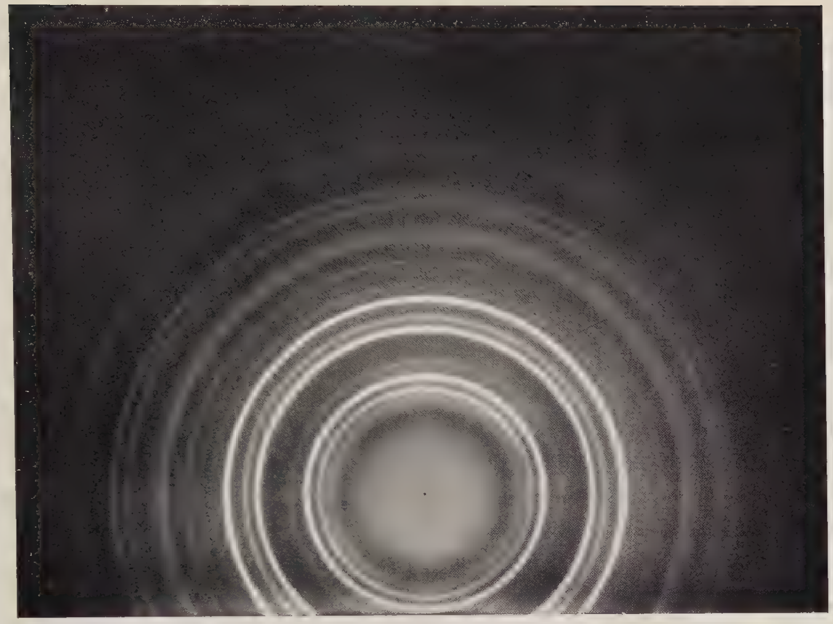


Figure 8. Transmission pattern from tin monoxide film.

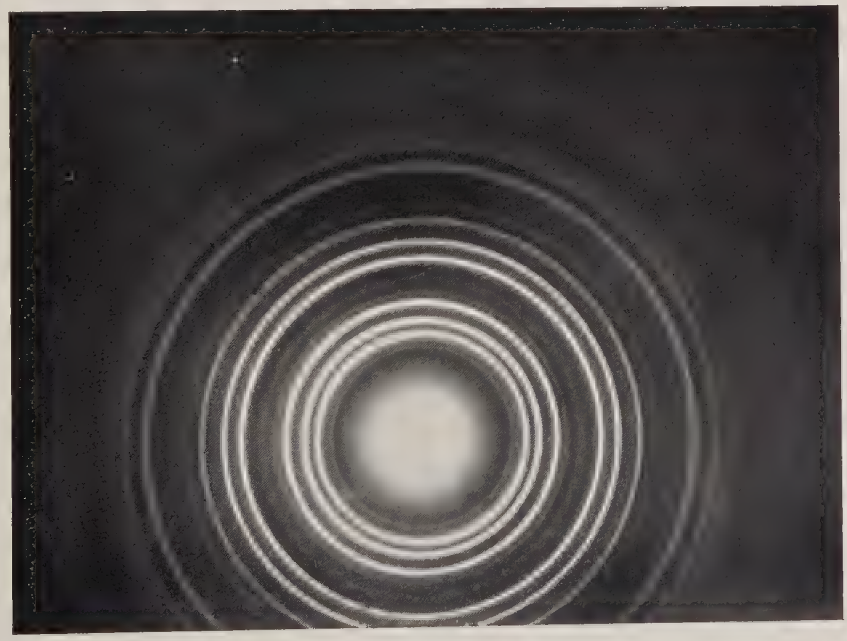


Figure 9. Transmission pattern from litharge film.

To face p. 120

oxide layer probably started as the monoxide and changed almost entirely to the dioxide owing to the higher temperature. Similar films were also made by Steinheil by heating tin foil. The relative intensities of the specimens made by both methods agreed well and showed little sign of any orientation, since all possible rings were present. The orientation was presumably destroyed on the transition from monoxide to dioxide. The rings are given in table 7 together with their intensities and lattice spacings.

Table 7

Indices of ring	Intensity of ring	Observed lattice spacing (A.)	Calculated lattice spacing (A.)
SnO (001)	Very weak	4·77	4·81
SnO (100)	Very weak	3·75	3·80
110	Very strong	3·39	3·34
Sn (200)	Very weak	2·90	2·92
101	Medium strong	2·65	2·63
200	Strong	2·39	2·36
210	Weak	2·15	2·12
211	Very strong	1·78	1·76
220	Weak	1·69	1·67
002	Weak	1·58	1·58
310	Medium strong	1·52	1·49
301	Very weak	1·35	1·33
202	Very weak	1·31	1·31
321	Medium strong	1·23	1·21
400	Medium	1·19	1·18
222	Weak	1·16	1·15
312	Medium strong	1·10	1·09
420	Weak	1·06	1·06
103	Medium weak	1·02	1·03
402, 123	Medium	0·96	0·96
510	Weak	0·93	0·93
332	Medium strong	0·92	0·92

Lead oxide. Films of lead oxide also were made with the same technique, and their diffraction patterns were photographed and analysed. These showed that the films consisted entirely of the rhombic form of PbO (litharge), the preferential orientation tending to set the crystals with their [001] planes in the plane of the film. This again caused rings due to the [00*z*] planes to be absent and those with the *z* index zero to appear with increased intensity. The rings observed by electron diffraction, figure 9, are given in table 8 together with their lattice spacings and intensities, and corresponding intensities obtained by Halla and Pawlek[(4)] with X-rays. A comparison of these intensities shows clearly the effect of the preferential orientation. In addition to the rings tabulated, two or three very faint rings were observed inside the [110] ring on some plates. These seemed to correspond to indices [010] [100] and $\frac{1}{2}$ [110], but may have some different origin such as an impurity with a larger unit cell.

Table 8

Indices of ring	Observed lattice spacing (A.)	Calculated lattice spacing (A.)	Observed electron-diffraction intensity	X-ray intensity
110	3·59	3·59	Medium weak	Weak
111	3·08	3·07	Strong	Strong
002	—	2·94	—	Very strong
200	2·73	2·75	Very strong	Medium
201	2·51	2·50	Very weak	Very weak
020	2·36	2·34	Very strong	Medium
112	2·21	2·28	Very weak	Very weak
202	2·03	2·02	Weak	Strong
003	—	1·96	—	Medium
022	1·87	1·85	Very weak	Strong
220	1·79	1·79	Very strong	Medium
113	1·72	1·72	Weak	Strong
311	1·65	1·64	Strong	Strong
203	—	1·60	—	Very weak
222	1·53	1·53	Weak	Strong
023	—	1·51	—	Weak
131	1·48	1·47	Strong	Strong
400	1·42	1·38	Medium weak	Very weak
114	1·37	1·36	Medium weak	Medium
223	—	1·31	—	Weak
204, 313	1·29	1·29	Weak	Strong
024, 402	1·25	1·25	Weak	Medium
133	—	1·21	—	Medium
330	1·18	1·19	Strong	Weak

§ 8. ACKNOWLEDGMENT

In conclusion, I should like to thank Prof. G. P. Thomson for his interest and many helpful suggestions in the course of the work.

REFERENCES

(1) Thomson and Fraser. *Proc. roy. Soc.* A, **128**, 641 (1930).
(2) Bethe. *Naturwissenschaften*, **15**, 787 (1927); **16**, 333 (1927).
(3) Kirchner and Raether. *Phys. Z.* **13**, 510 (1931).
Raether. *Z. f. Phys.* **78**, 527 (1932).
(4) Halla and Pawlek. *Z. f. Phys. Chem.* **128**, 49 (1927).
(5) Finch and Quarrell. *Proc. Phys. Soc.* **46**, 148 (1933).
(6) Ewald and Herrmann. *Structurbericht, Z. f. Krist.* Spec. (1931).
(7) Ponte. *Ann. de Physique*, **13**, 395 (1930).
(8) Bragg and Darbyshire. *Trans. Farad. Soc.* **28**, 522 (1932).
(9) Hull and Davey. *Phys. Rev.* **17**, 549 (1921).
(10) Steinheil. *Ann. der Phys.* **19**, 5, 472 (1934).

DISCUSSION

Lord Rayleigh mentioned the fact that when sodium-nitrate crystals are deposited on a calc-spar crystal, they try to orient themselves as nearly as possible in accordance with the orientation of the latter, although the difference in structure between the two substances prevents this result from being effected completely. In the present instance, when lead-oxide crystals were formed on the surface of a large single crystal of metal did the metal atoms of the oxide remain in the positions which they had occupied in the metallic state?

Prof. G. I. Finch. In view of the reputed purity of the zinc which Dr Quarrell and I had used, we had suggested that the faint extra inner ring in our oriented ZnO patterns might be a $\frac{1}{2}$ [102] of normal ZnO, but emphasized the need for caution by

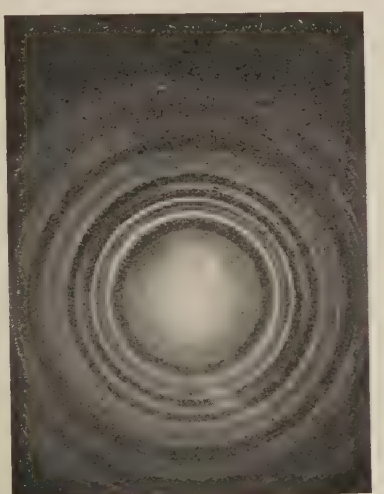


Figure A.

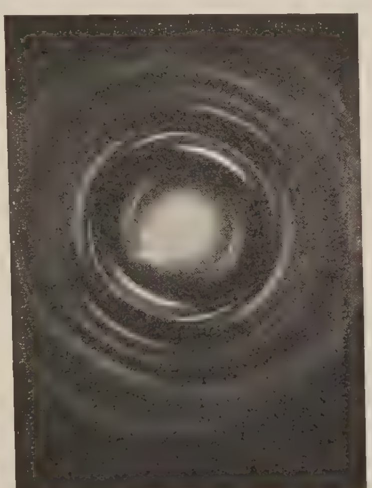


Figure B.

pointing out the fact that the supposed half-order and [102] rings arced in different directions. Since then Dr Wilman and I have found that this extra inner ring may be due to a trace of an impurity which can be almost wholly removed by skimming off the first few layers of oxide scum formed on the surface of a melt of a fresh batch of forensic zinc. The first skimming consists in the main of this impurity and is quite different in texture from a zinc-oxide scum. The transmission pattern obtained with a film of the impurity normal to the beam is shown in figure A. It has a complicated structure of a relatively low order of symmetry. One of the prominent inner rings corresponds to a spacing of 3·79 A., and this is the supposed half-order ring previously observed. Otherwise the pattern has nothing in common with either Zn, ZnO or pseudomorphic ZnO.

At first sight figure A, like many of our ZnO patterns and like those now discussed by the author, appears to be due to a random crystal array; inclination of the specimen plane to the beam by only 20° from the normal sufficed, however, to

yield an arced pattern, figure B, thus proving that the crystals tended to point in a common direction. The remarkable elliptical appearance of the arced pattern is, of course, due to an illusion.

Prof. G. P. Thomson. With reference to the single crystals of oxide which form on the surface of the lead when it solidifies, it seems to me that this is a different kind of phenomenon from those referred to by the President or Prof. Finch. In the author's experiments the small crystals of oxide are present throughout the experiment, even when the lead is melted, and the formation of a single crystal must be due to a rotation and orientation of these small pieces of crystal, brought about by the atomic forces in the metal at the moment when it solidifies. As far as I know this is a new effect, and it is of interest from the point of view of the study of crystallization.

I should also like to refer to the apparatus used by the author. This, as is stated in the paper, was designed by Mr C. G. Fraser of Aberdeen. In respect of the method of moving the specimen by means of a large coned ground joint carrying two smaller cones, figure 2, the apparatus has been imitated in others made at the Imperial College. Four such in all are now in use and some have been in operation for nearly three years. They have given very little trouble and I can recommend the mechanism as an efficient one for moving objects in a vacuum. There has been no difficulty arising from sticking of the large joints, of diameter up to 9 cm., in spite of the great pressure of the atmosphere on them.

Author's reply. In reply to Lord Rayleigh's question: the lead-oxide film on the surface of the large lead crystal was originally formed while the metal was liquid, and as the atoms in a liquid have no fixed positions there would be no relation between their positions in the lead and in the oxide. It is very probable, though the point was not investigated, that the single-crystal oxide film, caused by the rearrangement of the small oxide crystals when the underlying metal solidified, had some definite orientation with respect to the metal crystal.

Prof. G. P. Thomson has brought out the difference between the effects observed in the experiments under discussion, and phenomena observed when one substance crystallizes on another giving oriented single crystals or pseudomorphic crystals. The formation of the single-crystal lead-oxide films was caused by rotation of pre-existing small oxide crystals under the regular atomic field of the underlying metal crystal.

The ground joints to which Prof. Thomson refers have been in use for about $1\frac{1}{2}$ years without any leakage or regrinding, and with only a very occasional regreasing, except in the case of the two large flat joints which needed regreasing rather more often owing to their removal for access to the photographic plate and specimen.

Prof. Finch's new diffraction pattern from the first skim of molten zinc is very interesting as it explains the diffraction ring corresponding to a lattice spacing of 3·79 A. and bears out the observation that this and the extra line obtained in the reflection patterns from molten zinc might both be due to the same oriented impurity.

535.243

THE USE OF MICROPHOTOMETRIC METHODS IN DIVIDED-BEAM SPECTROPHOTOMETRY

By D. H. FOLLETT, M.A., A.Inst.P., Adam Hilger, Ltd.

Communicated by Dr F. Simeon, August 17, 1934. *Read November* 16, 1934.

ABSTRACT. A microphotometer is described in which two photocells, connected differentially, are used to indicate the difference in transmission of the individual spectra in the pairs produced in divided-beam methods of spectrophotometry. Match points are thus found with much greater ease than by the visual method; much of the uncertainty which arises in visual matching disappears, though some remains owing to the irregularities of the photographed plate.

§ 1. INTRODUCTION

The use of a microphotometer for finding the match points in divided-beam methods of spectrophotometry has been suggested but so far seems never to have been put into practice. The reason for this may be that apparatus designed specially for the purpose is necessary, since the use of the ordinary form of microphotometer would be very laborious and its adaptation to this special purpose is not practicable.

§ 2. CONSTRUCTION OF APPARATUS

Figure 1 shows diagrammatically the lay-out of an instrument designed for this purpose. L contains the light-source, a 6-volt 18-watt motor head-lamp running from batteries or the mains, whichever is the more convenient. The prism P deflects the beam at right angles on to a 25-mm. microscope objective M_1; this forms an image of the filament of the lamp on the plate, which is placed with its film side uppermost on the stage S. The linear dimensions of the image are one tenth of those of the filament. The second objective M_2 forms an image of the plate on the slit B which is in the centre of the screen A. The magnification of this system is $\times 10$. Since the filament is imaged on the plate, there is also a unit-magnification image of the filament on this slit. Each objective is furnished with a focusing motion. Two photoelectric cells of the rectifier type are mounted behind the screen, as described later. Between L and P is a lens, not shown in the diagram, which can be pushed into the beam as required; we shall refer to this as the *viewing-lens.*

The purpose of the viewing-lens is to simplify the setting of the plate on the stage. When this lens is in position, with a plate on the stage, a small patch of light, instead of an image of the filament, falls on the plate; and on the screen A appears a projected image of the spectrum under examination. The largest possible area

that can be imaged on the plate is a circular patch of the same dimensions as the aperture of the microscope objectives; the position and focal length of the viewing-lens was therefore chosen so that only this area was illuminated, the image on *A* being thus made as bright as possible. When a plate taken with some such instrument as the Spekker photometer is on the stage, an image of a pair of spectra separated only by a fine dividing-line appears on the screen. The position of the plate is adjusted until the image of the dividing-line falls exactly on a horizontal line engraved on the slit and passing through its centre; under this condition the light forming the image of one spectrum falls on one photocell, and that forming the other falls on another photocell. Figure 2 shows diagrammatically the optical arrangements whereby this is achieved.

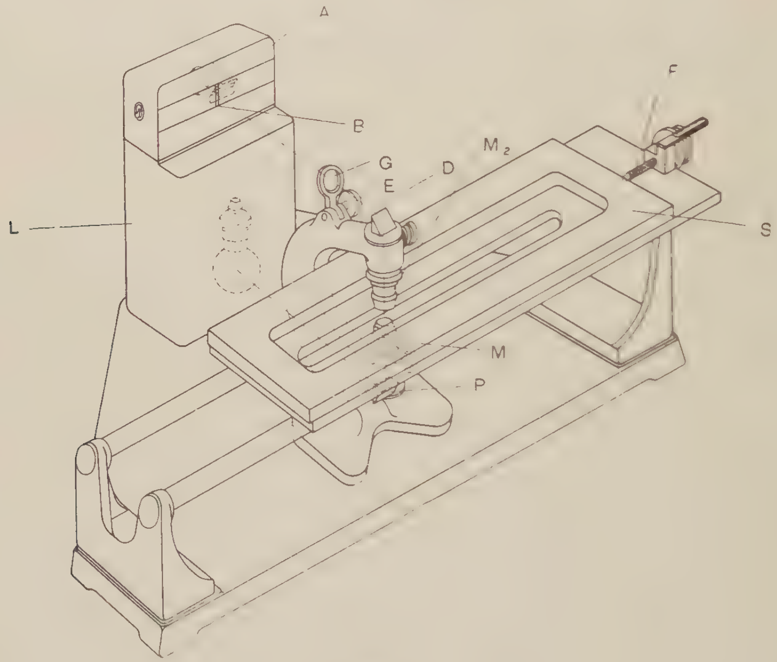
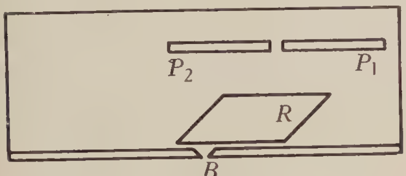


Figure 1.

The lower drawing shows an elevation, the upper a plan view of the photocell housing. The image of the dividing-line in the absorption spectrum falls on the straight line *H* which is engraved through the centre of the slit *B*, perpendicular to it. A rhomb *R* mounted behind the slit collects the light forming the image of one of the spectra and directs it on to the photocell P_1, while light forming the image of the other passes straight through on to the photocell P_2. The position of the rhomb is adjusted so that the projection of the line *H* falls along the centre line of a black band, 1 mm. wide, painted along the upper edge of the rhomb and indicated in the figure by the shaded area. This black band allows a little latitude in the positioning of the plate on the stage and prevents any difficulty over the finite width of the dividing-line.

The light falling on the cells must be divided so that the individual cells produce equal currents when an unexposed part of the plate is projected on the screen. The parallel glass plate G, figure 1, is provided for this adjustment; it is mounted so that its inclination to the beam can be varied between positions $22\frac{1}{2}°$ either side of the vertical; its thickness (about 3·5 mm.) is such that the image of the filament is displaced 1 mm. when the plate is moved from one extreme position to the other. Thus by moving the glass plate the proportion of the image of the filament falling on either side of the dividing-line can be adjusted until the amount of light falling on each cell is such as to produce equal currents in these cells.



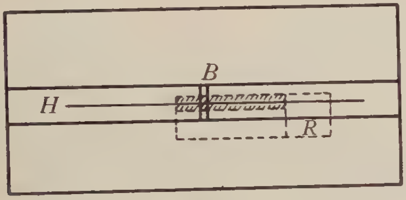


Figure 2.

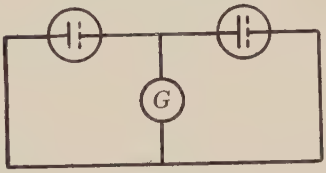


Figure 3.

The slit B is made with one fixed and one movable jaw; the movable one is held in position by a screw through a slotted hole, and for adjustment this screw is loosened and the jaw is slipped along in its slide. This is a sufficiently sensitive adjustment: the use of a screw is an unnecessary luxury, as the linear magnification of the image of the plate on the slit is $\times 10$, and therefore the slit-width can be about 10 times the slit-width of the spectrograph with which the absorption photograph was taken.

The photoelectric circuit is shown in figure 3. The photoelectric cells used are made by Electrocell G.m.b.H., Berlin. The light-sensitive element can be provided as an unmounted disc 25 mm. in diameter, and this form is convenient and suitable. These cells have a high sensitivity, about 400μA. per lumen. The matching of the intensity, current characteristics for the two cells was tested by putting an Ilford wedge on the stage of the instrument and using it exactly as if it were an absorption spectrum. The beams were balanced with the light passing through the thin end of the wedge; the wedge was traversed towards the thick end, and it was observed that the balance was nowhere disturbed. The galvanometer was the A.M. type of the Cambridge Instrument Co.; its period was 6 sec., its resistance was 450 Ω., and its sensitivity about 1000 mm./μA. with the lamp and scale two metres from the mirror.

It is a great advantage to be able to use the rectifier type of photocell, as its current is very easily measured with the galvanometer; the emission type would involve the use of an amplifier which is far less convenient than the galvanometer. In order to obtain the utmost sensitivity, however, it is necessary to form an image of the source on the slit, and since the spectrum cannot then be seen on the screen, the addition of the viewing-lens already described becomes necessary.

The sensitivity of the instrument is such that the deflection obtained through a clear plate on one beam only is about 240 mm. with a slit 0·3 mm. wide while the lamp is run at 7 volts. This slit is not unduly wide, for in absorption work the spectrograph slit may permissibly be wider than in emission work. The deflection obtained through a line may appear rather low in many cases but actually it is quite sufficient; any increase does not increase the accuracy of matching, for the limit of accuracy is imposed by irregularities of the plate as will be described later. The sensitivity is fully sufficient to reach this limit.

§ 3. METHOD OF USE

The procedure for finding match points is very simple. First with the plate on the stage the necessary focusing adjustments are made; the plate is then moved so that an unexposed part, in the neighbourhood of the match point, is imaged on the slit and then the plate *G* is adjusted until the galvanometer shows no deflection. Then the plate is placed so that the image of the dividing-line between the two spectra falls on the engraved line on the slit, and the stage is traversed until the deflection of the galvanometer becomes zero; the match point is then imaged on the slit in the screen. The stage can be traversed as a whole along the guides; in the neighbourhood of a match point it can be clamped and a more sensitive adjustment can be made by means of the screw *F*, which is 25 mm. long.

The wave-length corresponding to the point at which the match point occurs must be determined. In the case of line spectra the match point will in general fall between two lines; these lines can be marked directly on the plate, and their wave-lengths can afterwards be determined by reference either to a map or to the wave-length scale on the plate. But in the case of continuous spectra a different procedure must in general be adopted; the difficulty is that, although a mark could be made on the plate to indicate the position of the match point, reference to the usual wave-length scale at the top and bottom of the plate is not sufficiently accurate owing to the curvature which is impressed on large plates in most spectrographs. The difficulty is overcome, however, by arranging that a line spectrum shall be taken with each absorption exposure, to lie in juxtaposition with the twin absorption spectra. This spectrum will appear on the screen with the absorption strip under examination. When the match point has been found the reading of the screw is observed; then the plate is traversed until one of the lines in the line spectrum falls on a continuation of the slit and the screw reading is again read. The line can

e marked and later identified, and since the distance of the match point from this ne is known in millimetres, the wave-length of the match point can be determined y simple interpolation or by the application of a Hartmann formula, according to ne accuracy required.

§ 4. RESULTS

It was found that match points could not be determined with the precision that ad been expected. In the case of line spectra it had been anticipated that in eneral the match point would not fall definitely on a line but would be between wo lines. It had been expected that measurements on the background between the nes would not be of much value, but that some method of interpolation could be sefully applied.

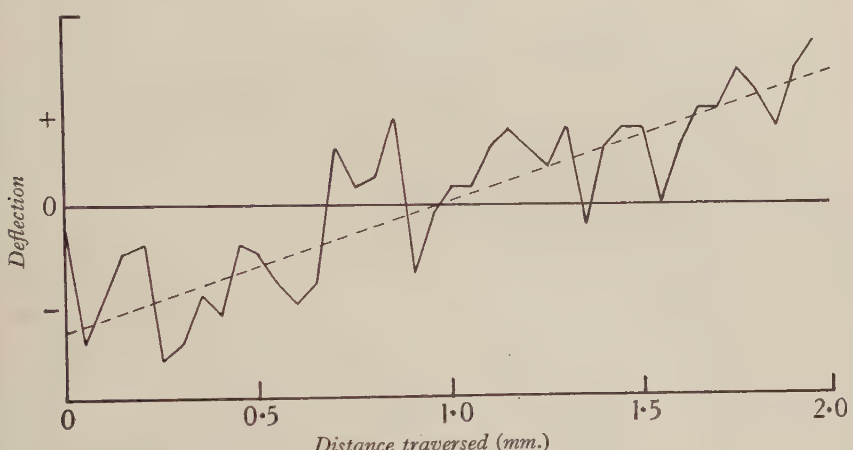


Figure 4.

It was found, however, that in cases where the absorption band was shallow, n equivocal kind of result was obtained, in which several successive lines gave leflections falling at random on either side of the zero; all that could be said in uch cases was that the match point lay somewhere in the midst of this group of ines. It was thought that this might be due either to pole effects arising from the ncorrect adjustment of the spark source in the photometer, or to irregularities in he photographic plate. Readings were therefore taken on plates obtained with a ontinuous spectrum on the Hilger-Nutting spectrophotometer, with a camera itted in place of the usual arrangements for visual readings. Any irregularities found here would presumably be due only to irregularities in the plate.

The actual deflection observed was taken at a large number of settings in the eighbourhood of match point. The area of each spectrum measured at a setting vas a rectangle measuring 0·15 × 0·05 mm. approximately, and the readings were aken at intervals of 0·05 mm. Figure 4 shows the result obtained with an Ilford uto-filter plate developed with metol-quinol. It will be observed that there is a

region nearly 0·9 mm. in length in which the line crosses and re-crosses the zero line. It is possible to draw a straight line which can be regarded as making an approximate fit with these points, and the point where this line crosses the zero line can be regarded as the match point. The match point can probably be determined fairly accurately in this way, but it is obviously out of the question for this procedure to be regarded as one to be generally adopted as it would be far too laborious.

The irregular nature of the line is presumably due to graininess of the emulsion, a factor which varies with both the developer and the type of plate. It is therefore clear that a combination of plate and developer showing the minimum of grain effect should be chosen. At the same time it is clear that the greater the slope of the line, the more clear-cut will become the match point. For a given absorption spectrum the slope will depend upon the contrast given by the plate; that is, it will increase as the slope γ of the plate-characteristic increases. The best combination of plate and developer for the work will be one with a small grain effect and a large γ. Experiments were therefore tried with different plates and developers.

In order to assess the relative suitability of the different plates it is desirable to derive for each plate some numerical value which can be regarded as a figure of merit. This was done in the following way. A straight line was drawn to fit the points obtained as nearly as possible, and the mean deviation of the points from
$\bar{D}$ this line was determined; this figure can be called $\bar{D}$. Let the slope of the line be
S called S. Then the value $S/\bar{D}$ was taken as the *figure of merit* for the plate in question. This figure is not intended to give a definite value for the merit of a particular plate; it is merely intended as a guide for arranging the plates in order of merit. Care was taken to see that the images measured on the different plates were all of approximately the same density; if this were not so the figure of merit as defined above would not be comparable for different plates. The results are given in the table.

Table

Plate	Developer	Figure of merit
Auto filter		0·17
Rapid process panchromatic	Potash-hydroquinone	0·34
	Glycine borax	0·41
		0·25
Panchromatic half-tone	Potash-hydroquinone	1·0
	Glycine	1·0

There seems to be no point in trying a further selection of plates. The panchromatic half-tone plate was chosen as it had a higher value of γ than any other plate we could find; the rapid process panchromatic was also chosen on account of its comparatively high γ. Plates with a lower γ would need to have an extremely low grain effect to compete with these, and since the grain effect with the panchromatic

half-tone plate seemed reasonably small it was concluded that a better plate was unlikely to be found among those with a lower γ.

The table shows that the panchromatic half-tone plates have the same figure of merit whether developed with glycine or potash-hydroquinone. Glycine is a developer which gives less grainy results than does potash-hydroquinone; but it also gives less contrast, and this disadvantage counterbalances the advantage of the reduced grain. Borax was found to have approximately the same grain as glycine but to give even less contrast.

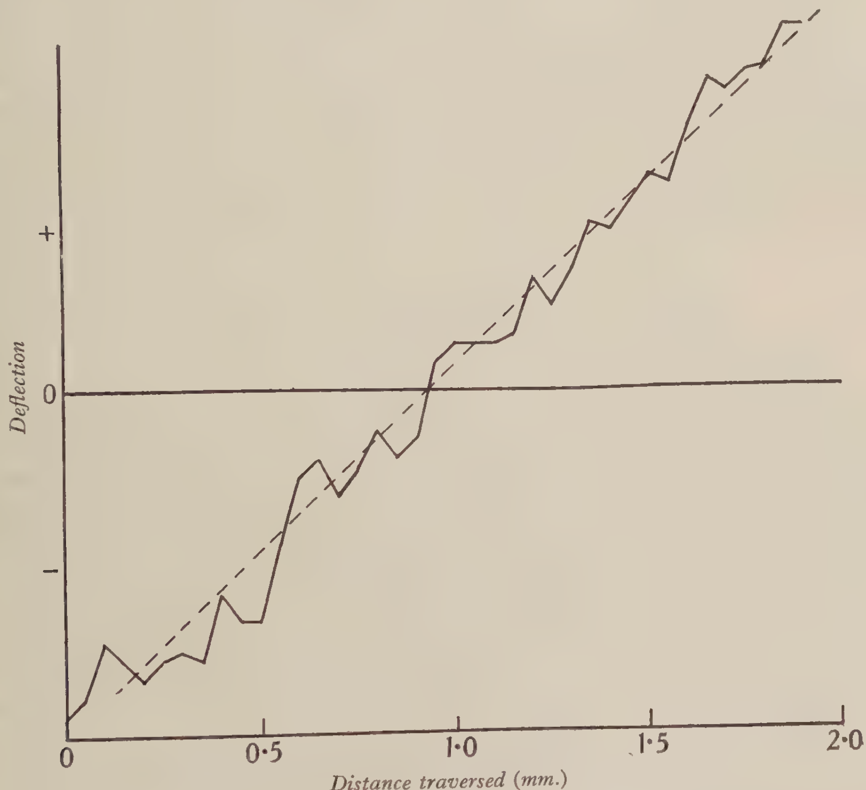


Figure 5.

We therefore conclude from these results that the best plate to use is the panchromatic half-tone; as developer potash-hydroquinone is preferable to glycine because it is more rapid in action. Figure 5 shows a plot of results obtained with this plate; it is directly comparable with figure 4, for the photometric conditions are the same and the density of the image at the match point is the same. The great superiority of the panchromatic half-tone plate is obvious.

The next stage is to determine what precision is actually obtainable in the determination of densities under ordinary conditions of use. With a coloured filter as specimen, a number of exposures were taken under the same photometric conditions and the match points were determined on the instrument. The procedure

was to watch the galvanometer deflection as the screw was turned and to observe the screw-reading when the match point was considered to be attained after the small fluctuations due to grain effect had been taken into consideration. A photograph of the copper arc was taken with each continuous spectrum, so that the two spectra were closely juxtaposed; a line in the copper spectrum then served as a reference point and the distance of the match point from this line was found by turning the screw until the projected image of the line fell on the slit in the screen.

Two sets of three spectra were examined; the only difference between the two sets was in respect of time of exposure, one set having about three times the transmission of the other at the match point. Two readings were taken for each spectrum, making twelve readings in all. These results were averaged; the mean deviation of the individual readings from this average was 0·05 mm. on the plate; the maximum deviation from the average was 0·11 mm.

In order to find what uncertainty these values correspond to in terms of the density of the specimen, a number of exposures were taken with the photometer circle set to different extinction values in the neighbourhood of that used in the above series of measurements, and the distances of the match points from one edge of the plate were measured. These values were plotted against the density D, and from the curve dD/dS was found in the region of the match point on which the above series of readings were taken. This value was found to be 0·1 2·5 where S is in millimetres. The mean deviation of 0·05 thus corresponds to an uncertainty of 0·002 in density, and the maximum deviation to just over 0·004 in density. This extreme value only occurred once; the nearest one to it was 0·08 mm., corresponding to about 0·003 in density, so that we can say that on a single reading there is a high probability that the uncertainty will be less than 0·003.

The above readings were taken at a wave-length of about 4000 A. The figure 0·003 for the sensitivity can only be taken to apply to that wave-length, since the value of γ varies with wave-length. However, this figure represents a sensitivity several times greater than has been claimed with visually spotted plates under the most favourable conditions(1); the sensitivity will be less in regions where γ is less, but as the sensitivity with visual spotting also decreases with γ the comparison between the two methods is not seriously modified by this circumstance. We can, therefore, make a general claim that the photoelectric method increases the accuracy of spotting several times over.

It does not follow from this that densities can be determined photographically with an accuracy of 0·003, for it has not been proved that the photometers in use are sufficiently accurate. They have been proved to be accurate within the limits which are imposed by visual spotting, but tests made with the more sensitive photoelectric method of matching may show that there is room for improvement in them. If this should prove to be the case, it does not follow that the higher accuracy of photoelectric spotting is wasted, for, in some applications of absorption spectrophotometry, sensitivity in detecting small differences in absorption is more important than accurate determination of absolute density.

When the light-source gives a line spectrum the sensitivity of matching is not

o good. The match point will in general be found to fall between two lines, which nay be separated by several angstroms. The match may be made on the background f it is sufficiently intense, but in general this will not be possible because the background is usually underexposed with the result that the contrast between the spectra is low. In the absence of the grain effect a fairly accurate interpolation between the two lines could be made, when the individual densities of the images are known. But the presence of the uncertainties due to grain effect reduces the accuracy obtainable. The accuracy may be still further reduced by the presence of the pole effect in the light-source; even with the best-designed instruments small uncertainties due to this effect are likely to appear.

We therefore conclude that photoelectric matching cannot be used to the best advantage when the light-source gives a line spectrum. But when it is used with a continuous spectrum a high accuracy in observing match points can be attained. Further, the operation is not so fatiguing as is visual spotting, although it is perhaps a little slower.

§ 5. USE AS ORDINARY MICROPHOTOMETER

The instrument is easily adapted for ordinary microphotometric measurements, which greatly increases its range of usefulness. This can be carried out in two ways. A double-pole double-throw switch is connected in the circuit in such a way that when it is put over the two cells become connected in parallel, instead of in the way shown in figure 3. The galvanometer then shows the sum of the currents due to the light passing through the upper and lower parts of the slit, instead of the difference between them. Alternatively, the photocell house can be removed and replaced by one containing a single photocell mounted directly behind the slit without any optical system; this gives a rather greater sensitivity than the first method, where nearly a third of the light from the filament is lost owing to the black band painted on the rhomb behind the slit.

The selection of single lines for measurement is obviously an easy matter; it is only necessary to adjust the position of the plates, with the viewing-lens in the beam, until the image of the line to be measured is seen to fall on the slit. As the amounts of light which fall on the photocell are very small, the current read by means of the galvanometer is very closely proportional to the light-intensity even though the galvanometer-resistance is as high as 450 Ω.

The sensitivity of the system is such that with the slit at 0·1 mm. the deflection obtained through the clear part of a plate is 220 mm. Its performance is indicated by figures 6 and 7. Figure 6 illustrates its definition. The object in this case was a piece of silvered glass with parallel lines cut on the silver. The lines were ruled 0·05 mm. apart and were approximately 0·025 mm. wide, so that the object consisted of a series of opaque and clear sections, each approximately 0·025 mm. wide. Figure 7 shows the result of a run on the 3100-A. triplet in the iron arc spectrum on a plate taken on the Hilger E 1 spectrograph.

Another application of this instrument is as a photomeasuring micrometer. Suppose one wishes to measure the distance apart of two lines; it is only necessary to set the plate on the stage so that the image of one of the lines falls on the slit and then to traverse the plate until the image of the other falls on the slit, and to read off from the screw the distance traversed. The accuracy of this result is limited by the accuracy of the screw, which is correct to 1/100 mm., but this is sufficient for nearly all purposes.

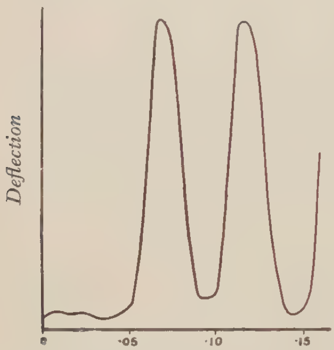


Figure 6. Engraved silvered glass plate.

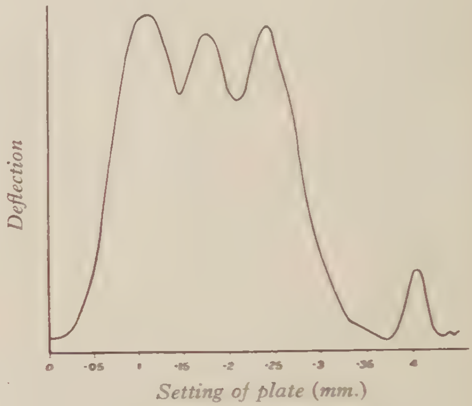


Figure 7. 3100 Fe triplet.

§ 6. ACKNOWLEDGMENTS

Acknowledgments are due to Messrs Adam Hilger Ltd., in whose laboratories this work was carried out, and at whose suggestion the results are published. I have had the benefit of useful discussions with Mr F. Twyman, F.R.S., and Dr F. Simeon, while Mr W. J. McCarthy and Mr R. A. C. Isbell have assisted ably with experimental work.

REFERENCE

(1) Twyman and Lothian. *Proc. phys. Soc.* **45**, 653 (1933).

DISCUSSION

Lord Rayleigh. I should like to ask the author why he uses colour-sensitive plates, which are in general more troublesome to manipulate? I have found that for photometric purposes plates coated on plate glass are much to be preferred, owing to the uniform thickness of the film.

Author's reply. Plates were selected by consideration of grain-size and contrast. Information concerning these points was obtained from certain data supplied by Messrs Ilford, some of which are reproduced in the paper by Twyman and Lothian

referred to. The question of colour-sensitivity was not considered at all; it just happened that on the list of plates for which data were available the two which appeared most promising were panchromatic. I do not think that, for any particular type of emulsion, the grain-size and contrast would be markedly different in a panchromatic and a non-panchromatic plate, but I have no definite information on this point; in the list of plates for which data were obtained from Messrs Ilford, panchromatic and non-panchromatic plates of the same emulsion type were not included.

I have had no actual experience of the use of plates coated on plate glass. It seems probable that their use would increase the precision of absorption-measurements made by this method.

534.22

ON THE BEHAVIOUR OF SUSPENDED PARTICLES IN AIR, AND THE VELOCITY OF SOUND AT SUPERSONIC FREQUENCIES

By E. B. PEARSON, B.Sc., Assistant Lecturer in Physics, University College, London

Communicated by Prof. E. N. da C. Andrade, August 2, 1934. Read November 16, 1934, by Prof. Andrade in the absence of the author.

ABSTRACT. In accordance with theory, particles of the size prevailing in ordinary cigarette smoke are found to act as obstacles in air vibrating at supersonic frequencies. This action leads to the formation of figures at the nodes in a resonance tube, from which measurements of the wave-length can be made. These, with a knowledge of the frequency of each piezo-electric crystal used to maintain the oscillations, enable values of the velocity of sound in air to be found at various frequencies in a range from 92·2 to 801·6 kc./sec. The values found show a definite dispersion of sound in this region, contrary to the results of previous experimenters with the Pierce acoustic interferometer. There are two maxima which are attributed to the separate effects of oxygen and nitrogen.

---

§ 1. INTRODUCTION

It has been shown by E. N. da C. Andrade[1] that any obstacle in vibrating air of sufficient amplitude is surrounded by a vortex system which leads to forces between two neighbouring obstacles. Two spheres whose line of centres is in the line of the vibration vector are in equilibrium at a certain distance apart: if the line of centres is transverse to the vibration vector the spheres are brought into contact. He points out that how far a particle acts as an obstacle, or is carried backwards and forwards by the medium, depends upon the size and mass of the particle and upon the frequency of vibration. For instance, cork-dust particles of the size used in his experiments take up about 96 per cent of the amplitude at 120 ~, and 50 per cent at 850 ~; but smoke particles, whose average diameter as determined by the rate of settling of a cloud was about 1 μ., take up about 99·9 per cent at 2000 ~, and so can be used as tracing-points at this frequency. It is pointed out, however, that at a frequency of 300000 kc./sec. smoke particles of radius 1 μ. should take up only 18 per cent of the motion of the air, and should thus act as obstacles. The formation of dust figures at sonic frequencies has been shown by Andrade[2] to be due to vortices around obstacles combined with a general circulation between node and antinode, which takes place in a vibrating column of gas enclosed in a tube. If this circulation takes place at supersonic frequencies smoke particles should therefore give rise to figures similar to some of the dust figures observed in a Kundt's tube. The experiments described in this paper were undertaken to in-

vestigate the behaviour of smoke particles in air vibrating at very high frequencies, with the particular object of seeing whether it was possible to obtain figures whereby the nodes or antinodes could be fixed and the velocity of sound deduced.

§ 2. DESCRIPTION OF APPARATUS

Piezo-electric quartz oscillators were used as the source of supersonic vibrations, because they maintain a very constant frequency, unaffected by temperature changes, and in addition are fairly easy to manipulate. The crystals used were supplied by Messrs Adam Hilger, and final frequency-measurements at the key frequencies were made by the National Physical Laboratory. The oscillating and sustaining circuit employed to excite the crystal is shown in figure 1; it was the standard Cady one, modified by the addition of a reaction coil. This was found

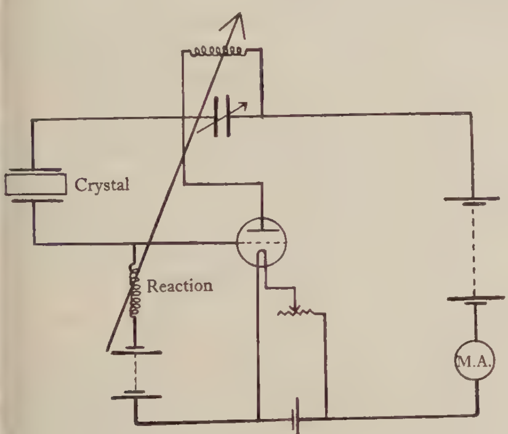


Figure 1.

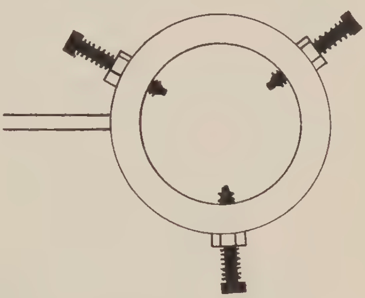


Figure 2.

necessary with the L.S. 5 valve used. The components were of the type used in standard wireless sets. To obtain adequate power a plate voltage of 360 was applied to the valve together with a negative grid bias of 20 volts, giving a plate current of 50 mA., the former being supplied by a battery of accumulators. The tuning and reaction coils were of the Igranic plug-in type and were mounted in a holder giving variable coupling, while variation of frequency was obtained by means of a Dubilier air condenser, the capacity of which was 0·0005 mF.

All the crystals were of the same diameter, 1·9 cm., and were cut with a shallow groove in the mid plane parallel to the ends. The method of mounting is shown in figure 2. A brass ring, 2·7 cm. in internal diameter, and 4 cm. in external diameter, carries three pointed screws which grip the crystal in the groove and are secured by locking nuts. The crystals were held loosely, as otherwise they do not oscillate freely. To the brass ring is screwed a rod, held in a stand so as to allow the position of the crystal to be adjusted.

The tube containing the air to be set in vibration was of internal diameter 2 cm., and was surrounded by a water-jacket to eliminate convection. The jacket

consisted of two plate-glass sides, measuring 53 cm. × 5 cm., cemented to a framework which had been cut from brass square-section tubing, and constituting the ends and base. This is shown in figure 3. The ends were brass plates pierced with circular holes, through which were soldered short lengths of brass tube marked *A*, *A* in the figure. The glass resonance tube passed through these tubes, at the points *B*, *B*, and water-tight joints were made with rubber collars of pressure tubing.

The frequency being constant for any given crystal, it is necessary to be able to vary the length of the resonant air column in order to make it contain a whole number of half-wave-lengths. At the very high frequencies, where the half-wave-length is of the order of 0·5 mm. and less, the adjustment of length is relatively unimportant, but for half-wave-lengths of 1 mm. and more it is essential. It was effected by means of a brass plunger, 2 cm. long, which fitted the tube as tightly as was compatible with easy movement and was soldered to the end of a length of 4-mm. brass tubing fitting loosely through the collar marked *C* in figure 3. A small hole was drilled through the centre of the plunger so that smoke could be introduced into the resonance space.

Figure 3.

As it was necessary to have an uninterrupted crystal surface for radiating the waves down the tubes, large electrodes could not be used. The front and back faces of the crystal were coppered, and small electrodes of copper foil soldered to springy wires made light contact with them. The electrode for the radiating face passed through a small hole, marked *D* in the figure, in the wall of the glass tube.

When a crystal was first tried out the correct sizes of tuning and reaction coils were found with the aid of a Sullivan heterodyne wave-meter, supplied with calibration curves for use with a series of plug-in inductance coils. From the nominal frequency of the crystal the wave-length of the electric waves in the oscillator circuit was calculated, and from the calibration curves the reading of the wave-meter appropriate to this frequency was found. The search coil connected to the wave-meter was fixed near the coils of the oscillator, whose tuning and reaction coils were changed until the heterodyne note was heard with a convenient setting of its condenser. This procedure was necessary because most crystals could be made to oscillate at several frequencies, by using sufficient reaction, but it is essential to use the fundamental mode, since that is the only one that can be maintained over a fairly wide range of capacity-change in the tuning circuit. The wave-meter was not sufficiently accurate for the purpose of the sound-velocity measurements described later, and the crystals were therefore sent to the National Physical Laboratory for accurate determination of the frequencies.

§ 3. ILLUMINATION OF THE TUBE

To render visible the movements of the smoke the image of a good optical lit, illuminated by an arc and condenser, was formed by means of a lens of focal ength 50 cm. and aperture 11 cm. The flat beam in the neighbourhood of the mage was used in various ways. For purely qualitative observations the whole ength of the tube was illuminated through a plate glass window closing the end, this window being pierced by a hole which allowed a glass rod to pass through it. A glass disc waxed to this rod acted as the reflector of the sound waves, instead of the brass plunger shown in figure 3, and allowed the length of the resonant column of air to be altered. The beam from the arc could thus be thrown down the axis of the tube and the phenomenon could be viewed as a whole.

This method of illumination was not intense enough when the spacing of the smoke figures was to be measured or when the figures were to be photographed, and for these purposes the following arrangement was used. The tube was set up perpendicular to the beam of light, and the image of the slit was formed along the axis of the tube through the surrounding water jacket. By this means a length of about 4 cm. of the inner walls of the resonance tube could be intensely illuminated. A screen of semicircular blackened metal placed underneath, in the water, served to cut off most of the high lights from the curved walls of the tube, and also helped to show up the illuminated portion.

§ 4. PHENOMENA OBSERVED

Preliminary experiments were made at a frequency of 92·2 kc./sec. with the longitudinal illumination already described. The approximate wave-length of the sound-waves in air at the temperature of the water in the bath was calculated, and the distance from crystal face to glass reflector was adjusted so as to contain a whole number of half-wave-lengths. The small electrode, which passed through the hole in the end of the tube, was arranged to be about 3 mm. away from the end of the crystal, while the other electrode just touched its back face. This disposition of the electrode was found to give best results with the crystals whose natural frequencies were 92, 147 and 220 kc./sec.; their relatively large vibrations would be damped if there was pressure on the surfaces. With crystals of higher frequency, however, both electrodes must touch the faces.

Smoke, dried by being blown through a long calcium-chloride tube, was then introduced into the length of tube between crystal and reflector, and the supersonic vibrations were started by varying the tuning condenser about the critical position found by the wave-meter method already described. The smoke was seen to circulate violently, and in a few seconds coagulation of the particles took place at equally spaced intervals, causing the formation of rings which extended up the walls of the tube and sometimes completely round them. When the oscillations were stopped the particles ceased to coagulate, and floated in the tube at random, but as soon as the vibrations were restarted the rings formed again. When the

crystal was kept oscillating for about half a minute a slight deposit was formed on the walls of the tube at the positions of the rings, and this remained after the oscillator had been switched off and the surplus smoke had been blown out of the tube.

Photographs were taken of the rings when they were in the process of formation, and also of the traces left on the walls of the tube. The earlier attempts were made with a Beck attachment fitted to the eyepiece of a microscope giving a magnification of about 6, by the method of longitudinal illumination, but insufficient light was reflected from the particles to give any record with exposures of 4 or 5 seconds, even when the fastest plates, 2000 H and D, were used. This length of exposure was the maximum possible when the rings were in process of formation, for they soon became indistinct owing to the deposition on the tube walls.

The transverse illumination, using slit and lens, caused too many high lights due to reflections from the outside of the tube, and consequently a square-section tube was made from four pieces of photographic negative glass, each 16 cm. long, cemented together at the edges to give an internal side of 2 cm. One end was closed with a brass plate, provided with a hole to allow a glass tube waxed to a square glass reflector to pass through it. By this means the length of resonant air column was varied. The other end was closed with a piece of card, with a round hole cut in it to allow the crystal to project inside. Sufficient light was obtained by using two identical optical systems of the type described, both being horizontal and making equal small angles with the normal to the tube. By this means photographs of the smoke figures seen during the oscillation of the air were taken with an exposure of 3 seconds. The result is shown in figure 4, where the line of the reflector forming the closed end is marked by an arrow. The curvature of the smoke planes was due to the strong circulation, which disturbs their alignment, but is only seen in the square tube. In a circular tube the planes are all quite flat and perpendicular to the axis. In all probability secondary disturbances in the air, due to the edges of the square tube, are responsible for the appearance of the planes in figure 4.

Examination of the traces left on the walls of the tube after oscillations had been stopped and surplus smoke had been blown out shows that they mark the nodes. Figure 5 is a photograph of these traces, formed in the circular tube at a frequency of 92·2 kc./sec. The picture was taken with the Beck attachment on a microscope, illumination being with the arc, slit and lens. Most of the high lights have been cut off by the blackened metal screen placed under the tube, but a few remain and are marked as such. This arrangement allows but a small width of the wall of the tube as being visible, but actually the traces extend completely round it. The sharp line marked by an arrow is the end of the brass reflector, and hence represents a node, so it is clear from measurements of the spacing of the traces that they themselves are at nodes. The double line forming each trace can be taken to represent the heaps of dust to either side of a node which, in a Kundt's dust tube, form the figures described by Andrade as "eyes". The double lines are shown more sharply in figure 6, which was taken at an early stage in the formation of the traces. As the amount of deposit increases the lines broaden so as to overlap, and at frequencies above 400 kc./sec. they are so close that the double structure is

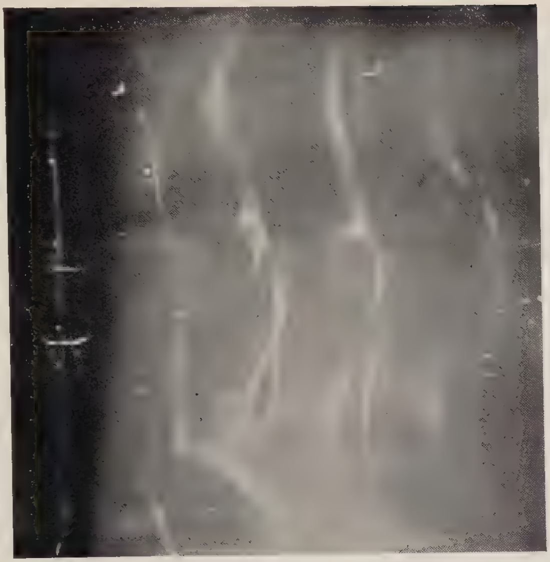


Figure 4.

High light due to reflection

Figure 5.

Figure 6.

[To face p. 140

arely seen. Before a fresh set of photographs was taken the tube was cleaned out by means of a wad of cotton wool moistened with methylated spirit, and then dried vith a soft cloth, pushed in on the end of a rod. The crystal was then replaced in oosition, and fresh smoke was blown in.

The phenomena here described can be explained on the view that smoke oarticles should act as obstacles in a gas vibrating at a very high frequency, and oehave as cork particles do at sonic frequencies, in the manner pointed out by Andrade[1]. Two particles which act as obstacles in a vibrating medium are brought and held together with their line of centres transverse to the vibration vector. It appears from the work of Whytlaw-Gray* that smoke-particles which come into contact as a result of ordinary casual impacts adhere to one another, so that it is to be expected that particles brought together by the forces generated by the vortex motion in the medium will likewise adhere. The effect will be more marked with the larger particles formed by the adhering of two of the original smoke particles, so the effect is cumulative, and rapid coagulation will take place to form the smoke planes seen when the supersonic vibrations are in progress. The fact that the planes disappear when oscillations cease shows that the coagulation is not of the ordinary type which is observed in a smoke cloud when this is left for a sufficient length of time; the coagulation must be due to minute vortices. When the oscillations are allowed to persist, the cumulative coagulation results in depositions, analogous to the eyes seen in a Kundt's tube, on the walls of the tube on either side of the nodes.

The phenomenon cannot be followed in all its details, as it was by Andrade in his work on dust figures formed at sonic frequencies, for with a crystal neither frequency nor intensity can be varied, and both of these factors enter into the mechanism of the formation of the various groupings and arrangements of particles in a Kundt's tube[2].

§ 5. DETERMINATION OF SOUND-VELOCITY

The distance between the mid lines of successive eyes deposited on the walls of the tube represents a half-wave-length $\frac{1}{2}\lambda_t$ of the sound in air, at the particular frequency of the crystal. This being accurately measured and the frequency being known from the report by the National Physical Laboratory on the crystal, a value of the velocity V_t at the temperature t° C. of the water-bath can be found from the relation

$$V_t = n\lambda_t,$$

where n is the frequency.

λ_t

V_t, t

n

The positions of forty or fifty nodes were noted by means of a Cambridge travelling microscope giving a magnification of about $\times 6$ and having graduations down to 0·01 mm. on a rotating head attached to the micrometer screw. Estimations to 0·002 mm. could easily be made, and readings were always taken from the closed end of the tube towards the crystal, but not nearer to the crystal than 6 cm. so that the region where the sound-velocity has been found by previous experimenters to increase slightly on account of proximity to the source was avoided.

* See e.g. Whytlaw-Gray and Patterson, *Smoke*.

The set of readings of nodal positions was divided into two equal groups, in the usual manner, and thus an accurate value of $\frac{1}{2}\lambda_t$ was found. The velocity at t° C.

V_0 was then evaluated and the velocity V_0 at 0° C. was deduced by means of the relation

$$V_0 = V_t - 0{\cdot}61\,t,$$

where V_0 and V_t are in metres per second, and t is in degrees centigrade.

At sonic frequencies the diameter of the tube influences the velocity, and the Kirchhoff correction must be applied in order to get the value of velocity in free space. This correction is expressed in the formula

$$V_T = V_A\{1 - 0{\cdot}54/2r\sqrt{(\pi n)}\},$$

V_T, V_A, r where V_T and V_A are the velocities in tube and free air respectively, and r is the radius of the tube in cm. In a tube of radius 1 cm., with a frequency of 92·2 kc./sec., which was the lowest used, the correction works out to be 1 in 2000, while at higher frequencies it becomes rapidly less. As the probable error in a series of values of V_0 came to about 1 in 800, the correction was neglected.

At the first frequency used, 92·2 kc./sec., a mean of eight values of V_0 gave a probable error of 0·4 m./sec. It was decided to measure the velocities at higher frequencies, and for this purpose eight more crystals, covering a range 127 to 1020 kc./sec., were employed. On repetition of the procedure already described it was found that similar phenomena were observed at each higher frequency, traces being left on the walls of the tube at the nodes, and these were measured with the microscope to evaluate $\frac{1}{2}\lambda_t$. Most of the crystals oscillated quite readily, and in some cases so violently that the crystal broke. Above 500 kc./sec., however, it was often a somewhat difficult matter to get oscillations sufficiently powerful to form the smoke figures, and careful variations of reaction coupling, filament current, and pressure of the electrodes on the crystal faces, were all necessary in order to bring about the desired result, except in the case of the 1020-kc./sec. crystal. In this instance no figures could be detected, although there was a fairly well-marked circulation in the smoke.

The value of V_0 obtained at each of the other frequencies was taken as the mean of eight determinations. A typical example of such a set is given in table 1.

Table 1

Frequency $n = 310{\cdot}198$ kc./sec.

Temperature (° C.)	Wave-length (mm.)	V_t (m./sec.)	V_0 (m./sec.)
16	1·104	342·5	332·8
16	1·106	343·1	333·2
16	1·103	342·1	332·4
16	1·104	342·5	332·8
17·5	1·108	343·7	333·0
18	1·112	344·9	333·8
15	1·104	342·5	333·3
17·5	1·109	344·0	333·3

Mean $V_0 = 333{\cdot}1 \pm 0{\cdot}2$ m./sec.

The values of V_0 obtained at each of the eight frequencies are given in table 2.

Table 2

n (kc./sec.)	V_0 (m./sec.)
92·20	330·7 ± 0·4
127·53	331·1 ± 0·4
219·28	332·6 ± 0·7
310·19	333·1 ± 0·2
356·78	331·9 ± 0·5
485·34	330·4 ± 0·5
628·96	331·1 ± 0·6
801·67	330·8 ± 0·6

The result of plotting V_0 against log n, which gives a better-spaced scale than n itself, is shown in figure 7.

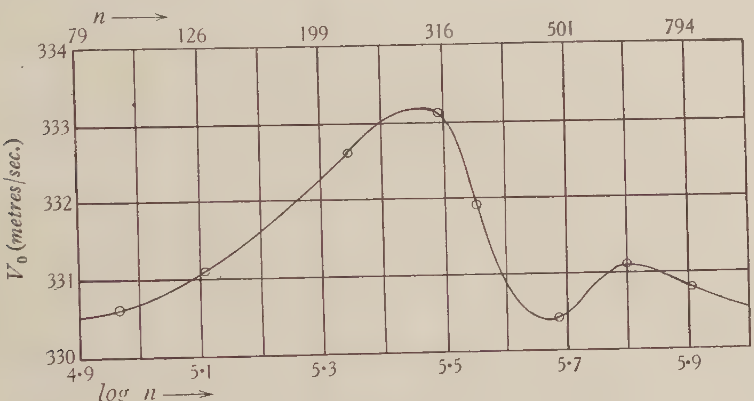


Figure 7.

§ 6. DISCUSSION OF RESULTS

The first measurements of V_0 were made at 92·2 and 310·2 kc./sec. and an increase above the normal value was found at 310·2. As air is a mixture, it was thought that each of the two principal gases might show its own maximum value of V_0 at an appropriate frequency, and as no further maximum appeared up to 485·3 kc./sec. on using the intermediate frequencies, the values above it were used, when the second small maximum was found at 630 kc./sec. The supposition that this is due to oxygen while the higher peak at 290 kc./sec. is due to nitrogen is further strengthened by the fact that the peak at 290 kc./sec. is almost exactly four times as large as that at 630 kc./sec.

A number of experimenters have worked on the problem of velocity-changes due to dispersion at supersonic frequencies, and although no recorded results show a marked change in air, it must be stressed that the values of V_0 at any given frequency, as found by different observers, differ appreciably amongst themselves.

With the exception of C. B. Vance[(3)], who used lycopodium to indicate the nodes, all other experimenters have employed the Pierce acoustic interferometer, in which the waves emitted from the oscillating crystal are reflected back on to it from a plane set parallel to it and capable of being moved perpendicular to its face by means of a micrometer screw. The difference in phase between reflected and emitted waves, which depends on the position of the reflector, causes the plate current of the oscillator to pass through a series of maxima, the distance moved through by the reflector between successive maxima being equal to one half-wave-length of the sound waves. The first important measurements of velocity in air were made by G. W. Pierce[(4)], who covered the range 90 to 1040 kc./sec. with four crystals having frequencies of 98·2, 205·6, 610·2 and 1034 kc./sec. At 610·2 kc./sec. appeared a very slight maximum value of V_0, 331·8 m./sec., to which Pierce himself attached no importance, since he considered it to be within the range of experimental error. His experiments were repeated in air[(5)] and later in oxygen[(6)] by W. H. Pielemeier, who detected no appreciable change of velocity in air at 303, 389 and 655 kc./sec., or in oxygen up to 316 kc./sec.

C. D. Reid[(7)] used the interferometer and detected a change in velocity with distance from the radiating face of the crystal, the value of V_0 being slightly higher when found close to it, and falling to a constant value at greater distances. The amount of change gets less with increasing frequency, being only 0·15 m./sec. at 140 kc./sec., the highest frequency he used. Up to that point he found no dispersion in air. The interferometer was used by P. T. Kao[(8)], who covered a range 40 to 1000 kc./sec., and also noticed slight increase in velocity quite near the source. He took the constant value of V_0 obtained at distances 30 to 40 cm., and this value of V_0 was constant at the various frequencies up to 1000 kc./sec., being 331·85 m./sec., correct to 0·1 per cent. This value is in fair agreement with Pierce's above 100 kc./sec., but the latter's values below this frequency are in excess of it.

The experiments of Vance[(3)] up to 200 kc./sec. gave a constant value of V_0, but his method seems open to criticism. Lycopodium powder was dusted into a tube fitted with a plunger in order to vary the length, and the open end was held close to the radiating face of the crystal oscillator. The powder was said to collect at the nodes and their spacing was then measured with a travelling microscope. Vance says that the nodes were so indistinctly marked that too great an error came in if the position of each was noted, and consequently only every tenth could be taken. This fact seems to indicate that the method is not adapted to accurate determinations of wave-length, and hence the figures for velocity cannot be very reliable. In any case, as the highest frequency used was only 200 kc./sec., the region of dispersion as found by the present writer had not been reached.

With carbon dioxide definite evidence of supersonic dispersion has been found. Pierce[(4)] obtained a rise in velocity of about 0·5 per cent between 98 and 205 kc./sec., but the work was incomplete and no measurements were made between 205 and 1034 kc./sec., at which latter frequency the gas was opaque to the sound and hence no figure for velocity was recorded. H. O. Kneser[(9)] used the interferometer method with carbon dioxide between 60 and 1480 kc./sec. His value for V_0 in the

region of 100 kc./sec. agrees fairly closely with that of Pierce, but above it he found a rapid increase of magnitude far larger than that detected by Pierce and reaching 4 per cent at 350 kc./sec. Above this frequency the velocity remained constant and did not fall to the normal value.

The results shown in the present paper thus differ very widely from those obtained for air in the past with the interferometer. The slightly higher values of V_0 that have been observed near the source cannot account for the increase found by the author at 290 kc./sec., and in any case the measurements of the spacing of the smoke-traces were never made near the crystal. Both the maximum at 290 kc./sec. and the minimum at 480 kc./sec. represent changes well above the experimental error, but no theoretical explanation can be offered to account for them. However, it seems a significant fact that these results, which are the first to show marked dispersion in air at frequencies above 100 kc./sec., have been obtained by a method which is the most direct of those so far used, depending as it does on a means of marking nodes with a material substance, exactly as has been done at sonic frequencies with cork dust. In addition, the length of the resonant air column is kept approximately constant during each determination, while in the case of the interferometer it is continually varied. As the supersonic waves attenuate fairly rapidly in air it is clear that as the reflector is moved closer to the course the reflected beam becomes more powerful and hence exerts a varying reaction on the crystal, so that it may well be that the maxima in the plate current do not always occur when the reflector moves through a distance equal to a half-wave-length, for the factors governing this change of current with phase-difference must be extremely complicated.

§ 7. ACKNOWLEDGMENT

Finally I wish to offer may sincerest thanks to Prof. Andrade, at whose instigation the problem was attempted, for his continual interest and invaluable suggestions during the course of the work, and also for the experimental facilities placed at my disposal.

REFERENCES

(1) *Proc. roy. Soc.* A, **134**, 445 (1931).
(2) *Philos. Trans.* A, **230**, 413 (1932).
(3) *Phys. Rev.* **39**, 737 (1932).
(4) *Proc. Amer. Acad. Arts Sci.* **60**, 284 (1925).
(5) *Phys. Rev.* **34**, 1184 (1929).
(6) *Phys. Rev.* **35**, 1005 (1930).
(7) *Phys. Rev.* **35**, 814 (1930).
(8) *C.R.* **193**, 21 (1931).
(9) *Ann. Phys.*, Lpz., **11**, 777 (1931).

DISCUSSION

Dr A. B. Wood. I am particularly interested in the latter portion of the paper, in which reference is made to the remarkable changes in the velocity of sound in air which occur at high frequencies. Like the author, I find it difficult to supply an

adequate theoretical explanation of the results. In such a case it is important to eliminate the possibilities of a defect in the experimental method. No doubt the author has already considered the possibility of gaseous impurities, such as water vapour or CO_2, being introduced into the tube with the smoke.

The statement on page 142 of the paper that some crystals vibrated so violently that they broke whilst others were difficult to excite into vibration of sufficient amplitudes, arouses the suspicion that at any rate some of the observed wave-velocity variation may be due to the differences in velocity-amplitude of the various crystals used in the measurements. It is a simple matter to show that the velocity-amplitude $\dot{\xi}$ at the antinodal faces of a half-wave-length resonator is given by

$$\dot{\xi}=\frac{P}{E}C,$$

where P is the stress at the nodal plane, E the elastic modulus and C the velocity of sound in the oscillator. The upper limit of P is set by the breakdown stress of the oscillator. Tables give various values for quartz, an average being $0{\cdot}5\times10^{10}$ dyne cm.$^{-2}$ If $C=5{\cdot}5\times10^5$ cm./sec. and $E=8\times10^{11}$ dyne cm.$^{-2}$, this gives the following value for the maximum velocity-amplitude of the face of the crystal:

$$\dot{\xi}_{\text{max.}}=3500 \text{ cm./sec.}$$

That is, the quartz crystals used in the experiments might have any velocity-amplitude up to 35 m./sec., resulting in a pressure-amplitude of 0·14 of an atmosphere in the air near the face of the quartz. To this must be added the possibility of a further increase of velocity-amplitude due to resonance in the air column contained in the tube. Such large-amplitude particle velocities might be expected to produce serious distortion of wave-form and increased wave velocity.

In a resonant air column of the kind used in the experiments it is unlikely that much decrease of velocity with distance from the crystal-face would be observed. The antinodes in the air column will, in a reasonable length of tube, have approximately the same velocity-amplitude, unless the gas produces strong attenuation. The author's remark on page 145 that "measurements of the spacing of the smoke-traces were never made near the crystal" does not, therefore, eliminate the possibility of high wave-velocities, due to resonance of the crystal and air-column.

It appears to be desirable that further experiments should be made with the 310-kc. crystal, which gives an abnormally high velocity, excited at different voltages or used in different gases. Such experiments would reveal how much of the observed change in sound-velocity is due to variation in velocity-amplitude of the vibration and how much to the nature of the gas.

I do not wish to suggest that the observed velocity-changes at high frequencies are not due to the causes mentioned in the paper, but I think that other possibilities such as those I have mentioned should be eliminated before a satisfactory theory can be established.

Prof. E. N. da C. Andrade. I am much impressed by Dr Wood's calculations, which are possibly not, however, as conclusive as he believes. The crystals which break do not crack across the nodal planes but at one edge, with cracks that appear

to be rather parallel to than normal to the axis of the cylinder. I am inclined, from examination of the crystals, to think that the fracture is largely due to marked weakening consequent on the cutting of the groove for holding the crystal. The effect of any minute sharp edges left on such an entrant cut is well known. Incidentally, the value given for the tensile strength of quartz in the *International Critical Tables**, quoted as "maximum observed", is about one-fifth† of that quoted by Dr Wood, and is presumably for a static load, and it further seems likely that, as only a few crystals failed, they were rather of minimum than of maximum strength. Lastly, I am in some little difficulty to see how the effect contemplated by Dr Wood could be reasonably expected to lead to a rise to a maximum followed by a drop, in the consistent way shown by all results.

As regards Kneser's theory, while, as I pointed out in reading the paper in the author's unavoidable absence, it is undoubtedly attractive and has had good success in some directions, Kneser's experiments themselves do not furnish any evidence of a maintenance of the high velocity at higher frequencies, but merely show that it is just attained. A subsequent drop is quite consistent with his experimental figures. There is, further, a difficulty on the question of absorption. Kneser himself alludes to Pierce's finding that CO_2 is practically opaque to sound waves in the region of 2×10^5 c./sec. and says that the differences of absorption are greater than is to be anticipated from the theory. He adds that the experimental result speaks rather for the occurrence of selective absorption in the region of dispersion, of which the theory can give no account. I also note that Sherratt and Griffiths, in their recent paper, while saying that they do not attach much weight to the discrepancy, nevertheless point out that with the gas, carbon monoxide, used by them "no agreement was found between the resulting sound-absorption in the gas and that to be expected from Kneser's theory". I do not think that the theoretical position is sufficiently well established to throw doubt on the experimental results of the author.

Mr J. H. Awbery. In presenting the paper, Prof. Andrade referred to the work of Kneser, whose theoretical curve of velocity as a function of frequency does not show a maximum, but rises gradually from one constant value to another. He pointed out that Kneser's observations in CO_2 did not extend far enough to verify the later part of the curve, i.e. to show that the course of the curve after the rise is horizontal. It may be of interest to note that in a paper read to the Royal Society on November 15, Dr Griffiths and Mr Sherratt provide indirect evidence of the truth of Kneser's theory. They worked at two frequencies only, but by using the theoretical formula for the curve given by Kneser, they deduce the true velocity of sound, and hence the ratio of the specific heats of the gas that they used (carbon monoxide). Their results up to 1800° C. agree with those deduced for specific heats from spectroscopic data, thus removing a discrepancy which had for some time been troublesome. This result at least makes it probable that the theory given by Kneser has in fact traced

* *International Critical Tables*, **4**, 22.
† Viz. $0{\cdot}91 \times 10^9$ dyne/cm. Voigt gives $1{\cdot}26 \times 10^9$ dyne/cm.

the true cause of the discrepancy, and that his theoretical curve would therefore be of the correct form.

I wonder if any tests were made to ascertain whether the frequency of the crystal when coupled to the resonating air column is the same as when it was calibrated—i.e. presumably with a very different acoustic load. Even a slight alteration in frequency would suffice to affect the shape of figure 7 materially.

Mr G. G. Sherratt. Figure 7 appears to me to be remarkable in that there is a maximum in the (velocity, frequency) curve. Since the work of Kneser and the consequent explanation of the discrepancy that has existed for a long time between spectroscopic specific heats and those determined from sound-velocity data, one would not be surprised to see an increase of velocity with frequency. But that velocity should subsequently decrease as the frequency is raised demands a further explanation. Kneser's work was founded on experimental evidence and has since received much experimental and theoretical support. Furthermore, it provides a reasonable physical picture to account for the increase of velocity with frequency.

If there is an increase over the frequency-range covered by the present experiments, it must presumably be due to the apparent disappearance of part of the rotational heat-capacity of the oxygen and nitrogen molecules. The small vibrational specific heat possessed by oxygen at room-temperatures disappears at much lower frequencies.

With regard to the query at the end of the paper as to the accuracy of the acoustic interferometer, it may be as well to state that the interferometer yields accurate half-wave-lengths if two precautions are observed. The first is that the sound-intensity must be sufficiently small for the limiting velocity to be obtained. The second is that the frequency must be independent of reflector-position.

In this connection it would be of interest to know whether the author took the precaution of verifying that the frequencies of the crystals were unaffected by variations in acoustic load.

Author's reply. In reply to Dr Wood: the only crystals that broke were those of the lower frequencies, before the velocity increased very much, which seems to indicate that the maximum amplitudes, which might be responsible for increased wave velocity, occurred at frequencies that gave almost normal velocity-values. As to gaseous impurities—the water vapour was removed, while any CO_2 would be expected to prevent a peak in the curve, for Kneser's experiments showed that the high velocity-values persisted as the frequency increased, in the case of that gas. Further experiments with different gases are already being considered.

In reply to Messrs Awbery and Sherratt: no tests were made to ascertain whether acoustic load influenced frequency, but I suggest that the frequency of a quartz crystal, which depends solely on its physical constants, would not be likely to be altered by the back pressure of the resonant air column. Kneser's figures for pure CO_2 will be checked in due course, when the apparatus for using gases other than air is built.

534.22

THE VELOCITY OF SOUND IN SHEET MATERIALS

By A. B. WOOD, D.Sc., F.Inst.P.,

AND

F. D. SMITH, D.Sc., A.M.I.E.E.,

Admiralty Research Laboratory, Teddington

Received March 16, 1934. *Read October* 19, 1934.

ABSTRACT. Chladni's figures have been produced at frequencies up to 20,000 by means of magnetostrictively excited tubes or rods of nickel. The velocity of sound in sheets of brass, iron, lead, ebonite, celluloid, cardboard and other materials has been measured with considerable accuracy by this means.

In a recent issue of *Nature** a note by R. C. Colwell refers to the excitation of Chladni's figures at high frequencies by means of a nickel rod excited magnetostrictively and arranged to set circular and square brass plates in vibration. The paper concludes with a statement that the plates "cannot take up such a high vibration as 15,000 per second."

As a matter of fact this method of exciting sheets of various materials into transverse vibration has been in use in this laboratory for a considerable time, particularly in connection with magnetostriction depth-sounding apparatus. The results obtained prove not only that thin sheets of metal or almost any solid material can vibrate transversely at these high frequencies, but also that measurements of the spacing of the nodes yield a tolerably good value for the velocity of sound appropriate to the thin metal sheet.

If a horizontal sheet of the material is sprinkled with sand and touched at a suitable spot with a tube or rod in vigorous high-frequency vibration, transverse waves travel over the sheet and are reflected back on reaching the edge or edges. The sheet is assumed to be large compared with a wave-length of the waves transversing it. The edge-reflected waves interfere with the direct waves and produce a stationary wave pattern in the sand. A resonating nickel tube has been found to give good results, eddy-current losses at these high frequencies being much less in a thin-walled tube than in a solid rod. With metallic sheets such as brass the stationary wave pattern is often complex, owing to interference of the direct wave with the waves reflected from the four edges of the sheet.

With ebonite, celluloid and other non-metallic sheets, however, the attenuation of the transverse wave is sufficient to leave only the simple pattern produced by the direct wave and that reflected from the nearest edge of the sheet; the waves reflected from the remoter edges are generally of insufficient amplitude to disturb the sand.

* *Nature*, Lond., **133**, 258 (1934)

If the sheet is circular, and is touched by the nickel tube at the centre, a series of equi-spaced sand-rings are produced from which the wave-length of the transverse vibration is easily measured. In all cases the sheet must be supported as lightly as possible so as to introduce a minimum of interference with the free vibration.

Thin nickel tubes have been used at resonance frequencies of 20,000 or more c./sec. with success, but it is advisable to select the frequency and the thickness of the sheet in accordance with the conditions mentioned below.

If the point of excitation is suitably chosen it is possible to obtain nodal lines running parallel to one edge of the plate, the measurement of the half-wave-length (distance apart of the nodal lines in the sand) being made in a direction normal to the free edge which reflects the waves. In such a case we may, with a fair degree of accuracy, write*

$$C=\frac{(1+k^2\kappa^2)^{\frac{1}{2}}}{k\kappa}C_t \qquad \text{......(1),}$$

C_t, C k, λ, κ

where C_t is the velocity of transverse waves, C is the velocity of sound appropriate to the sheet, $k=2\pi/\lambda$, where λ is the transverse wave-length, and κ is the radius of gyration of the cross-section of the sheet about the neutral axis of bending. When $k\kappa$ is small this reduces approximately to

$$C=C_t/k\kappa \qquad \text{......(2).}$$

The accuracy of the velocity determined by this method is dependent ultimately on the accuracy with which the thickness t of the sheet and the transverse wave-length can be measured, since $\kappa=t/\sqrt{12}$ and $k=2\pi/\lambda$.

Photographs of sand figures on sheets of brass, celluloid and ebonite are shown in figures 1, 2 and 3. The relevant data for these are given in the table at nos. 5, 16 and 14. Even a relatively soft metal like lead gives very clear nodal lines, and very thin metal sheets, such as 0·004-in. stalloy (no. 7 in the table), give good results.

The method is shown at its best when large sheets of relatively non-resonant materials such as celluloid, cardboard or ebonite are used. In these cases the simple pattern obtained by reflection from the edge nearest to the source agrees remarkably well with the calculated distribution of the nodal lines. The latter are produced by the interference between the direct waves from the source and the reflected waves from the image source, and consist of a family of hyperbolae. An example of this is illustrated in the photograph (figure 3). The principal hyperbolic nodal lines are clearly defined, whilst a second family of hyperbolic nodal lines, due to reflection from the remoter opposite edge (not included in the photograph), are faintly shown crossing the principal series.

The table gives the values of the velocity of sound in various sheet materials determined by this method.

* See Lamb, *Sound*, equation (12), p. 123 (1910).

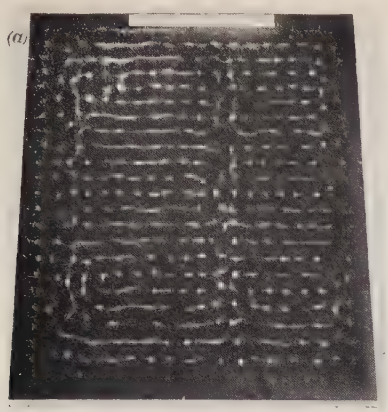


Figure 1. Brass 18 in. square and $\frac{1}{8}$ in. thick.

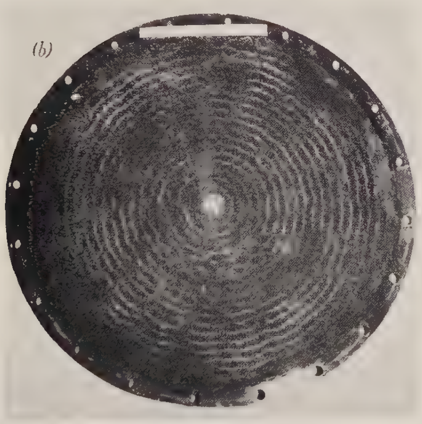
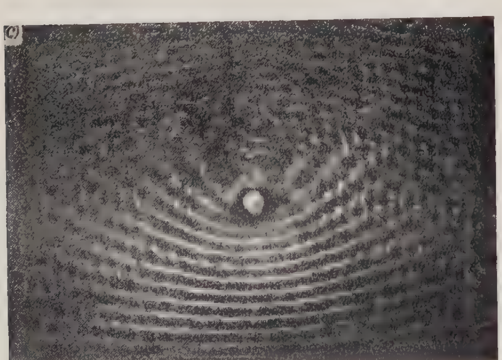


Figure 2. Celluloid 20 in. in diameter and $\frac{1}{16}$ in. thick.



Reflecting edge

Figure 3. Ebonite 3 ft. × $18\frac{1}{2}$ in. × $\frac{1}{8}$ in.

[*To face p.* 150

Table I

No.	Sheet material	Thickness (in.)	No. of half-wave-lengths counted	Total length (cm.)	λ (cm.)	N (c./sec.)	C_t measured (cm./sec.)	$k\kappa$	C (cm./sec.)
1	Brass	0·019	42	40·3	1·92	9,070	$17·4 \times 10^3$	0·0455	$3·82 \times 10^5$
2	,,	0·019	36	30	1·67	12,100	20·2	0·0521	3·88
3	,,	0·019	40	27	1·35	18,400	24·9	0·0645	3·86
4	,,	0·128	10	25	5·0	9,140	45·6	0·118	3·88
5	,,	0·128	16	27·5	3·56	18,300	65·2	0·165	4·0
6	Iron	0·095	22	55	5·0	9,500	47·5	0·087	5·46
7	Stalloy	0·004	10	5·4	1·08	9,070	9·8	0·017	5·8
8	Aluminium	0·067	12	26·5	4·42	9,500	42·0	0·0695	6·05
9	,,	0·103	10	27·5	5·5	9,500	52·2	0·086	6·1
10	Nickel	0·125	10	29·4	5·88	9,500	55·8	0·0978	5·71
11	Alpax (cast)*	0·166	12	28·5	4·75	18,260	86·6	0·160	5·5
12	Ebonite	0·046	12	8·6	1·43	18,290	26·2	0·147	1·73
13	,,	0·046	9	9·1	2·02	9,070	18·3	0·104	1·77
14	,,	0·124	10	11·5	2·3	18,200	41·9	0·249	1·73
15	,,	0·124	18	30·0	3·33	9,070	30·2	0·172	1·78
16	Celluloid (clear)	0·062	12	14·1	2·35	9,070	21·5	0·121	1·79
17	Lorival†	0·127	12	12·8	2·1	18,200	38·3	0·278	1·44
18	,,	0·260	13	28	4·3	9,070	38·9	0·277	1·45
19	Cardboard	0·070	8	10·6	2·65	9,070	24·0	0·121	1·98
20	Lead	0·010	12	5·4	0·9	9,070	8·16	0·051	1·59
21	,,	0·040	10	8·9	1·78	9,070	16·1	0·103	1·56
22	,,	0·060	10	10·7	2·14	9,070	19·4	0·129	1·51
23	,,	0·070	8	9·55	2·4	9,070	21·8	0·134	1·63
24	Keramot†	0·030	12	6·8	1·13	18,250	20·7	0·122	1·7

* Al-Si alloy. † Commercial insulating materials.

It will generally be found, where comparisons are possible, that the values of the velocities of sound in the sheet materials tabulated above are intermediate between the values obtained for the longitudinal velocities in rods and in bulk. The velocities in rod, sheet and bulk are the uni-, bi- and tri-dimensional cases in which the velocities are given by

$$\sqrt{\frac{E}{\rho}}, \quad \sqrt{\frac{E}{\rho\,(1-\sigma^2)}}, \quad \text{and} \quad \sqrt{\frac{E\,(1-\sigma)}{\rho\,(1-\sigma-2\sigma^2)}}$$

respectively, where ρ is the density, E Young's modulus, and σ Poisson's ratio.

DISCUSSION

Dr E. J. Irons. The measurement of velocities in sheet materials is a subject of some interest to me as it has, during the past session, engaged the attention of Mr Huffington and myself at East London College. Our aim has been to determine the velocity of a pulse produced in a material by displacements at various angles to its plane, and our purpose that of developing a method to provide constants that we hope may ultimately prove of service in moving-coil loud-speaker design. We

have employed a direct timing method and hope to publish in due course an account of these and other experiments made with our apparatus.

I suggest that a reference to the source of equation (1) would render the reading of the paper easier.

Dr N. W. McLachlan. Have the authors tried their method with paper from 8 to 10 mils thick? The velocity of sound in cardboard given in the table is of the same order as the value I obtained for paper some years ago by calculation from direct measurements of elasticity and density*.

AUTHORS' reply. We shall look forward with interest to seeing Dr Irons's results. The reference for equation (1) has now been inserted in the paper. In reply to Dr McLachlan: the method as applied to thin cardboard is equally applicable to thin paper. Even tissue paper 5×10^{-4} in. thick gives clear nodal lines from which the velocity may be calculated. As we pointed out in the paper, however, the accuracy ultimately depends on the accuracy with which the thickness of the sheet is measured.

* *Phil. Mag.* **13**, 115 (1932).

534·323

THE MEASUREMENT OF LOUDNESS

By N. R. CAMPBELL, Sc.D., F.Inst.P.,

AND

G. C. MARRIS, B.Sc., M.I.E.E.

A communication from the staff of the Research Laboratories of The General Electric Company Limited, Wembley.

ABSTRACT. No new scale of loudness is proposed, nor are those already proposed examined completely. The object of the paper is to examine proposed scales from one point of view only, and to determine only whether any one of these is distinguished from the remainder by being founded more firmly on facts and being more free from arbitrary convention. We are not concerned primarily with such questions as convenience or practicable accuracy.

In § 1 the facts (*a*)–(*f*) implied by all adequate scales of loudness are stated in a manner that does not (like most statements) imply that any scale is possible. It is concluded that these facts are not sufficient to distinguish from each other the infinite number of scales that might be or have been proposed. If one scale is to be distinguished from another by facts, some other facts must be introduced which are not so intimately involved in the meaning of loudness. There is an important difference in this respect between scales for tones of the same frequency and scales for all sounds.

In § 2 proposed scales for a single frequency are considered under the following headings: (i) mental estimates; (ii) equal relations; (iii) just-perceptible differences; (iv) thresholds.

The presumed facts (*g*)–(*k*) are involved in these scales and in part determine them; but the facts as known at present are not sufficient to force on us the selection of one particular scale, though they may force us to reject some.

In § 3 the problem of sounds in general is considered. Here again it is concluded that the facts, though they may limit further the choice of scales, do not determine one scale uniquely. On the other hand, if a scale for a single frequency could be agreed upon, work on the lines of Fletcher and Munson's latest publication would solve the problems peculiar to complex tones so far as it is soluble at all.

In § 4 some more general questions are treated briefly. It is urged that the arbitrary element in all scales may be advantageous as well as disadvantageous, for it permits the use of different scales for different purposes. To the question whether a scale uniquely determined by facts can ever be set up, we reply with a guarded negative; such a scale is conceivable, but all evidence available at present indicates that it is no more than conceivable.

§ 1. THE FACTS IMPLIED BY SCALES OF LOUDNESS

The following discussion is as little abstruse and as closely confined to the immediate question as possible. But some reference to the general theory of measurement cannot be avoided. In our use of terms (such as *A* and *B* magnitudes, regular and irregular errors, constants, accuracy and sensitivity) not defined in the text, but not unambiguous in current usage, we have followed Campbell's *Measurement and Calculation*, to which reference may be made.

The facts of audition are sometimes stated in such a way as to imply that an entirely satisfactory scale of loudness is already available. Since our object is to enquire whether such a scale is possible, this method of statement is likely to introduce confusion. We will start therefore by re-stating certain familiar facts in a way free from objection for our purpose. These facts are as follows:

Fact (a)

(*a*) If two sounds are sounded alternately at suitable intervals, any observer who can hear both can judge whether they are equally loud or, if not, which is the louder. Such judgments define what is meant by loudness; no statement about loudness in what follows has any meaning except in reference to these judgments. In virtue of them the observer can arrange in an order of loudness all sounds which can be thus sounded alternately.

Fact (b)

(*b*) If the same pair of sounds is concerned, whether pure tones or complex, any observer is consistent with himself at successive trials of (*a*), apart from irregularities of the kind that can be attributed to a finite sensitivity. If the pair are pure tones of the same frequency between the upper and lower thresholds of the observers, different observers agree amongst themselves to the same degree, though they may have different sensitivities. If the sounds are not pure tones of the same frequency, observers may differ regularly; older observers tend to find high notes less loud relatively to low notes than younger observers.

Fact (c)

(*c*) Equally loud pure tones of the same frequency are equivalent in respect of all kinds of audition. Other equally loud sounds are not always equivalent. Thus of two equally loud sounds one may be more effective than the other in masking a third sound. Loudness, if it were defined with reference to simultaneous sounds, would not be the same thing as loudness defined (as we have defined it) with reference to successive sounds.

Fact (d)

(*d*) Loudness is not additive. The operation most nearly satisfying the conditions for addition is simultaneous sounding. But in order that it should satisfy them, the following relations must be true:

If $\qquad X=X'$, then $X+Y=X'+Y \qquad$ ......(1).

If ϵ is such that

$\epsilon+X>X$ for any X, then $\epsilon+X>X$ for all X's $\qquad$ ......(2).

Relation (1) holds if X, X' are pure tones of the same frequency, but not always otherwise. Relation (2) never holds; ϵ can always be chosen so that if it is sounded simultaneously with a comparatively faint sound the combination is louder than that sound, but that if it is sounded simultaneously with a loud sound, the combination is no louder than that sound.

The foregoing facts can be stated without reference to any magnitude or measurable property. The next two establish a relation between loudness on the one hand and, on the other, frequency and intensity, which are physical magnitudes measurable by processes universally accepted.

Fact (e)
k, p_k, x_k

(*e*) Any sound can be resolved (e.g. by Fourier analysis) into n components, of which any one k has the frequency p_k and the intensity x_k. The loudness of a sound is "determined by" these (p_k, x_k)'s. By that we mean (the definition must be noted)

hat two sounds having *all* their (p_k, x_k)'s the same are equally loud. On the other and, sounds not having all their (p_k, x_k)'s the same may also be equally loud. The oudness of a sound generally, if not universally, increases with any x_k when the emaining x_k's are constant.

(*f*) Among pure tones of the same frequency, loudness is determined (in the bove sense) by their intensity only. The order of loudness is the order of the ntensity. Fact (*f*) may be regarded as a special case of fact (*e*); but it is convenient o state it separately. *Fact (f)*

If we had an accepted "scale of loudness," we could assign a numeral X to epresent the loudness of any sound. In virtue of (*e*) it would then be possible to ind some relation of the form

$$X = a . F(p_k, x_k) \ldots \quad (k = 1 \text{ to } n) \qquad \ldots\ldots(3)$$

X, F

valid for all sounds and, in virtue of (*f*), a relation

$$X = a . F_p(x) \qquad \ldots\ldots(4)$$

F_p, x

valid for all pure tones of the same frequency p. p

F must be single-valued, having the same value when all (p_k, x_k)'s are the same; F_p must be single-valued and monotonic. The problem of finding F and F_p would be in principle one of mere experimenting. It might turn out that F and F_p were not any known analytic functions; the relations might have to be expressed by graphs or mere tables.

All the scales of loudness that have been proposed in effect propose a function F or F_p such that the loudness of a sound can be measured (that is to say, a numeral X assigned to it) by means of a relation (3) or (4). One test that they must pass, if they are to be acceptable, is that when the measured arguments (p_k, x_k) for complex sounds are inserted in the proposed F, or the measured argument x for pure tones of a single frequency into the proposed F_p, then X resulting from (3) or (4) must actually be the same for all equally loud sounds.

If "equally loud" means "equally loud for all observers," no F can possibly pass the test; for when some sounds not of the same frequency are compared, all observers do not agree whether they are equally loud. The most that we can hope for is to find an F which will satisfy the test for some concordant group of observers. But for such a group the test is a real one, which some scales may satisfy and some not. The requirement that F shall have the same value for certain different sets of (p_k, x_k) imposes a very severe restriction on our choice of F. Thus it is agreed that the decibel-above-threshold scale does *not* pass the test for any important group; while the scale recently proposed by Fletcher and Munson* apparently does pass it within a certain range.

On the other hand not only can scales pass the test for F_p, but all proposed scales actually do so. Since there is only one variable, the only restriction imposed by the test is that F_p should be single-valued and monotonic. It would be satisfied if we simply put $F_p(x) = x$, and made our scale of loudness simply the scale of

* H. Fletcher and W. A. Munson, *J. acoust. Soc. Amer.* 5, 82 (1933).

intensity. But everyone agrees that such a scale of loudness is not acceptable, even for tones of a single frequency. There must be some other test or tests which scales have to pass and our first business is to find out what they are.

One test may be that F_ν must be the function to which an acceptable F reduces when there is only one component (p_k, x_k). One reason why the decibel scale is not acceptable, even for a single frequency, is doubtless that it does not pass this test. But this cannot be the only test. If it were, it would clearly be useless to try to establish a scale for a single frequency before we had a scale valid for different frequencies; yet there have been many attempts to do so. Again it seems clear that even our test for F is not sufficient. It is to be observed that if any F passes our test, so must any single-valued and monotonic function of F; yet all such functions of F cannot be equally acceptable. Whatever form of F we obtained in the first instance, it seems likely that *some* such function of it could be found such that, for a single component, it reduced simply to x. That function at least would not be acceptable.

This point is of great importance and must be illustrated by reference to the scales of physical magnitudes. The same magnitude can sometimes have several different scales, each of which is a single-valued function of another. Thus the transmission S and the photographic density δ of a plate, or the decay constant λ and the period T of a radioactive element, are the same magnitudes measured on different scales, related in the same way as any pair of proposed scales of loudness for a single frequency. Since $\delta = -\log_{10} S$ and $T = 1/\lambda$, the translation of a statement expressed in one scale into a statement expressed in the other is a mere algebraical transformation having no physical significance. It is a matter of mere convenience which we use. This is sometimes overlooked, because one form always is the more convenient; we are never tempted to use another.

On the other hand there are other magnitudes in respect of which it is not indifferent which of two scales related in this manner we use. The chief of these (possibly the only ones) are those which are additive and can be measured as A magnitudes. Thus it is not indifferent whether we write Ohm's law $E = a.I$ or $E^2 = b.I^2$; for the magnitude a (which is measured by this law as a B magnitude) is additive and measurable as an A magnitude also; but b is not. In this case there is a very definite reason for adopting one "scale" of resistance rather than another; one expresses, the other conceals, the supremely important fact of additiveness.

Now what is our reason for preferring one scale of loudness to another? Why are we quite sure that loudness ought not to be measured on a scale of intensity, but prepared to consider that it might be measured on the decibel scale by the logarithm of the intensity? Any statement in one scale can be translated by simple algebra into a statement in the other; and it does not seem that mere mathematical convenience is definite enough to make us so sure. There must here also be some facts revealed by one scale and concealed by another. It cannot be that which determines our choice in respect of resistances; for loudness is not additive. What is this fact? Is there indeed a real fact?

That is the main question discussed in this paper. We shall try to answer it by examining, under the headings set forth in the abstract, the scales that have been

roposed. We shall ask what are the facts on which the scales are based and whether ney really do provide an adequate reason for preferring one scale to another. n § 2 scales for pure tones of a single frequency will alone be considered; but the uffix to F_p will be dropped for brevity. Certain facts (*g*), (*h*), (*j*), (*k*), which aparently underlie these scales, will be assumed without the expression of any pinion whether they really are facts. Our question is only—If they are facts, do hey determine a scale?

First perhaps we ought to say what we mean by a fact. We shrink from atempting any precise definition; but it will probably be admitted that the distinction etween a fact and an arbitrary convention lies in its validity for all observers. In liscussing sounds of different frequencies, it would perhaps be too rigid to insist hat nothing valid for some observers and not for others is a fact; for, in virtue of 'fact'' (*b*), which is part of the definition of loudness, there could then be no facts nd no scale determined by them. But in considering tones of a single frequency listant from a threshold, agreement in (*a*) and (*b*) is universal and it is permissible o be strict; since there are facts valid for all observers, we may demand that any ther "facts" should be equally universal.

§ 2. TONES OF A SINGLE FREQUENCY

(i) *Mental estimates*

Fact (*g*). If observers are told that a certain weak sound is to be represented by 1 and a certain strong sound by 100, they can be induced with a little persuasion to assign numerals by simple "guessing" to other sounds, especially those of intermediate loudness. *Fact (g)*

If their assignments were as instinctive and as concordant as those concerned in fact (*a*) they would deserve as much attention; they would lead at once to a relation (4). But then we should never have realized that there was any difficulty in measuring loudness. We might need a convention concerning the numerals to be assigned to the fixed weak and loud sounds; but if those could be fixed, no further question would arise; nobody would have any doubt what numeral was to be attached to any other sound. But the guesses are not as instinctive as judgments of equal loudness. Advocates of the method of mental estimates forget that the need for advocacy is the chief bar to its success.

Nor are they concordant. Superimposed on the irregularities of a single observer at successive trials are undoubted regular differences between different observers. Moreover it is at least highly probable that estimates are affected by the order of presentation of the sounds, and that a lower numeral would be assigned to a sound presented after a long series of sounds all louder than itself than to the same sound if presented after a long series all weaker than itself. From such discordant judgments a scale universally acceptable can be derived only by arbitrary conventions.

But though this method of obtaining a scale of loudness has never been considered seriously, it needs a little further examination, because it involves a question

concerned also in our next method which is of much greater importance. How do we come to assign numerals at all by mere guessing and without actual measurement? We all constantly do it in every-day life; we talk of a 50-per-cent chance and of being twice as happy, when we have not carried through any process remotely resembling the measurement of any physical magnitude.

Surely the answer is that we are using some analogy between the circumstances of the judgment and those in which true measurement is possible. Thus at one extreme of this process of guessing is the "estimation of tenths." In such estimations we are undoubtedly using our familiarity with scales truly divided by actual measurement; if we had never seen such scales, the process would be impossible. Now, if we are using any analogy, the question arises whether the analogy is sound and whether the analogous facts are true. If there were some definite reason why the measurement of length broke down below the limit to which scales are divided, or if actually divided scales were entirely arbitrary and not based on true measurement, we should not attach much importance to estimation by eye. If that is so, before we accept mental estimates as true in the sense that the results of measurement are true, we should discover the analogy on which they are based and examine its validity very carefully. Universal agreement is no test of truth at all; for the most frequent source of universal error in the history of mankind has been an unquestioning acceptance of false analogies. And that is precisely what is most to be feared here.

(ii) *Equal relations*

It appears that, in addition to the judgments of equality of loudness concerned in fact (*a*), similar judgments can be made concerning equality of relations between pairs of sounds that are not equally loud. Expressed more accurately, the fact that we now add to our list is:

Fact (*h*) (*h*) Given any pair of sounds U', V' of the same frequency and any other sound X, an observer can find a sound Y whose loudness is related to the loudness of X as the loudness of V' is related to the loudness of U'. X may, of course, be either U' or V'. All observers will agree in selecting the same Y within the limits fixed by their sensitivity. This cannot always be true when X is near the upper or lower limit of audibility; but it will suffice if it is true over a considerable range.

In order to discover how this fact may be used to limit further our choice of F, let us leave loudness for a moment and consider pitch, which is determined by frequency as loudness is by intensity*. Fact (*h*) *mutatis mutandis* is true of pitch. A and B define a musical interval; X and Y are tones separated by that interval. If we plot the frequency y of Y against the frequency x of X, we shall actually obtain a straight line,

a

$$y=a.x \qquad \text{......(5)}.$$

If we took another pair of sounds U'', V'', defining another musical interval, and repeated the process, we should obtain another straight line and relation (5), but

* Pitch differs from loudness in that equality of pitch (which defines what we mean by pitch) can be determined by simultaneous sounding as well as by alternate sounding; but this is irrelevant.

h a different value for a. Accordingly (5) sums up all the observations we could ke in the matter in a single compendious formula.

Now these are precisely the circumstances in which, if we were making physical easurements (e.g. Ohm's law), we should say that we had found a new "derived" ignitude (e.g. resistance). It is therefore reasonable to say that we have found a w magnitude, pitch-interval, which is measured by a.

These facts impose a restriction on the forms of F in a relation $X=F(x)$, by iich it is permissible to represent the relation between pitch X and frequency x. ey show that our new magnitude a, pitch-interval, is determined (in the sense fined above) by the pitches of two sounds, the frequencies of which are related the same straight line (5). That is to say, if we assume a relation $X=F(x)$, we ust have

$$a=\phi(X, Y) \qquad \text{......(6)},$$ ϕ

here ϕ is a single-valued function and Y is the pitch of the sound whose frequency related to the frequency of the sound of pitch X by (5). But this frequency is ax. onsequently we must have

$$a=\phi\{F(x), F(ax)\} \qquad \text{......(7)}.$$

(7) gives us a relation between the functions ϕ and F, but it still does not de- rmine them uniquely. Thus we might have

$$F(z)\equiv Az^n; \quad \phi(u, v)\equiv(v/u)^{1/n} \qquad \text{......(8)},$$

here A and n are any real finite constants,

r

$$F(z)\equiv \log_n x; \quad \phi(u, v)\equiv n^{(v-u)} \qquad \text{......(9)},$$

here n is any real finite constant.

Now let us resume our discussion of loudness. We can again plot y against x x, y being now intensities) for a given "interval of loudness" defined by A, B. But the graphs of y against x are not now straight lines; they are not any known lgebraical curves. The graphs for different intervals of loudness, though similar, lo not clearly differ from each other simply by the variation of a single parameter a. We cannot therefore measure intervals of loudness as we can measure intervals of pitch; and we cannot summarize all our observations in a single compendious formula, analogous to the equation $y=ax$ for pitch. We can represent them only by family S of curves, each for a single interval.

We are therefore in an even worse position for choosing a scale of loudness than for choosing a scale of pitch; for the condition expressed by (7) is lacking. But we can try to regain condition (7) by assuming that the family of curves do represent a single function with a single varying parameter, although we cannot for the present determine it algebraically. This assumption (which will be called A) may be expressed by *Assumption*

$$y=S(a, x) \qquad \text{......(10)},$$

where S is an unknown single-valued and probably monotonic function. The S

fact that the interval of loudness between two sounds is determined by a relati between their intensities is now expressed, not by (7) but by

$$a=\phi[F\{x\},\ F\{S(a,x)\}] \qquad \text{......(11).}$$

Before we pass on, we must note that assumption A is precarious. We oft discover empirically a set of relations that can be expressed by a family of curv and sometimes later find the algebraic expression of those relations. It does not th always turn out that the curves differ in a *single* parameter; two or more paramete may be involved, which are not single-valued functions of each other*. If this shou be the truth here, our subsequent procedure is unjustified. For if (10) should reall be written $y=S(a, b, c, \dots x)$, (11) no longer follows. However it will appear th the assumption will, in some measure, provide its own test.

But (11) does not in general determine F uniquely, any more than (7), even whe S is known. In general, there will again be an infinite number of pairs of function ϕ, F, satisfying (11). The form of S, known graphically, may exclude certai alternatives; thus, since the curves S do not now pass through the origin, solution of the form (8) are (almost certainly) excluded from the start; but a new infinity o alternatives becomes open. The usual procedure at this stage is to assume a form for ϕ. If we make such an assumption we can now determine F graphically, usin the graphically known S; the process need not be explained, because it has ofte been carried out and is familiar. If successful, it results in a graph $X=F(x)$ givin a relation between loudness and intensity, such that pairs of points on it whos ordinates are related in some manner determined by our choice of ϕ have the sam interval of loudness and lie on a single curve of the original family. All the curve of that family are then summed up in this single graph; we have achieved the sam unification as we achieved for pitch by means of the formula $y=ax$.

But it is to be noticed that the process might not be successful. It might turn out that if we used one of the original family of curves S, one graph F resulted, but that, if we used another, another graph F resulted. This possibility has generally been recognized; it has been realized that a scale of loudness determined by equal relations is legitimate only if all the curves S give the same graph F. If they do, the facts confirm assumption A, which is involved in the process.

If they do not, the procedure is unjustified unless there is one member of the family S so clearly distinguished from the rest that it is reasonable to select this

S_0 member and neglect the rest. If this member is S_0, we can then replace (11) by

$$F\{S_0(x)\}=\phi[F\{x\}] \qquad \text{......(12);}$$

again, if we choose ϕ we can determine F. Recent work† seems to suggest that there is such a unique relation, somewhat analogous to the octave in the field of pitch; just as untrained ears can identify an octave more definitely and certainly than (say) a tone, so they can identify one particular relation between loudness more definitely

Fact (j) and certainly than any other. This fact will be called fact (j).

* But it sometimes *does* turn out. Stefan's law and Planck's law were both discovered empirically as families of curves before their algebraic expression by functions with a single parameter was found.

† B. G. Churcher, A. J. King and H. Davies, *J. Instn. elect. Engrs.*, 75, 401 (1934).

But whether we use fact (h) or fact (j), F cannot be determined without assuming the form of ϕ. Those* who have used fact (h) alone have assumed

$$\phi(u, v) \equiv v - u \qquad \text{......(13)},$$

which means that the loudnesses of equally related sounds are to be represented by numerals with a constant arithmetical difference. For example, the sound of loudness 100 must be related to that of loudness 90 as the sound of loudness 60 is to that of loudness 50. This assumption, which will be called B, is also implied in many discussions of the decibel scale; thus this scale has been criticized on the ground that the sounds represented on it by 100, 90, 60, 50 are *not* related in the manner just stated.

Assumption

But it cannot be insisted too strongly that the assumption is not self-evident; it needs justification. The scale based on it will be determined by facts only if this justification consists in facts. Since nobody seems to have discussed the matter, we do not know what reasons would generally be alleged. But we can suggest three which might be alleged; it is to be noticed that, of the following reasons, (α) is based on facts, (β) is neutral with respect to facts, (γ) is contrary to facts. Here they are:

(α) It might turn out that, if any other form of ϕ were assumed, assumption A would not be confirmed; a different curve F would be obtained from different members of the original family S. Whether it does turn out so is a question of fact that does not seem to have been decided. Owing to the limited precision of the observations, it would clearly be impossible to prove that a slightly different form of ϕ would destroy assumption A; but if it could be proved that no form differing greatly from that defined by $\phi \equiv v - u$ is consistent with assumption A, there would be some reason for assumption B. But it would not be a very strong reason; for there is no *a priori* evidence for either A or B; it might be a mere chance that the errors in assumptions A and B compensated each other. If, on the other hand—as will probably prove the fact—other forms of ϕ are consistent with assumption A, there is evidence for A but none at all for B.

(β) Assumption B is more convenient in one way than any other. No other leads to a curve for F in which sounds standing in the same relation of loudness are represented by ordinates differing by equal lengths; it is much easier to pick out from a graph ordinates differing by equal lengths than ordinates related in any other manner, e.g. having lengths in the same ratio. But this, though highly convenient, is no more than convenient. Its convenience arises solely from the peculiarities of graphical representation, which cannot have any bearing on the facts of audition. So long as we are content to have an arbitrary element in our "scale of loudness" we are entitled to take advantage of this convenience; but we must not forget its presence and be surprised if other scales, based on other facts, do not agree with it.

(γ) What we have called an *interval* of loudness is sometimes called a *difference* of loudness; and pairs of sounds which lie on the same curve of the family S are said to show an equal difference of loudness. It is thence concluded that, if loudness is

* We believe that this assumption has been made, but we cannot now trace the reference. All that is important to our argument is that it might be made; and this is clearly implied by the familiar criticism of the decibel scale, mentioned in the text.

to be measured at all, such pairs of sounds must be represented by numerals showing the same arithmetical difference. This is a pure blunder*, due to a confusion between the wide meaning of "difference" in ordinary discourse (when it means any relation of which non-identity is an essential element) and its narrow meaning in arithmetic (when it implies far more than mere non-identity). If "interval" is substituted for "difference," the fallacy appears at once; it becomes obvious that an interval may be represented by a constant ratio (as in the Pythagorean treatment of pitch) as well as by a constant arithmetical difference.

Indeed there is a very definite reason for *not* identifying interval with arithmetical difference in both cases, and for *not* assuming the relation $\phi \equiv v - u$. The reason is that there is a very important class of magnitudes that are characterized by a relation for which arithmetical difference is the only appropriate numerical representation. These are the additive magnitudes, such as length, mass, and electrical resistance and capacity. The class and the relation are supremely important because, in virtue of the relation, magnitudes of this class and no others are measurable fundamentally —that is to say, without any other magnitude being measured first. If loudness were additive, none of the questions we are discussing would arise; there would be no question how it is to be measured; the only doubts remaining would concern "units" or (as we prefer to say) factors and standards. Assumption *A* may be permissible even if loudness is not additive, just as it is permissible for pitch, which is not even approximately additive. But it is unfortunate to select among the infinity of possible ϕ's that one which, if it were specially appropriate and distinguished from all the others, would make any selection unnecessary and the whole problem nugatory.

Assumption C

As we have noted, another possible assumption, which will be termed *C*, is that

$$\phi(u, v) \equiv v/u \qquad \text{......(14)},$$

so that the loudnesses of equally related sounds are to be represented by numerals in a constant ratio. Before we discuss it, we must insist that it is definitely inconsistent with assumption *B*; equal relations cannot be represented both by equal differences and by equal ratios; if we adopt *C* we must cease to expect equality in the relations of loudness 100 to loudness 90 and of loudness 60 to loudness 50. *B* and *C* can be true at the same time only if there are two distinct kinds of relations between loudness, one of which is appropriately associated with *B* and another with *C*. It is extremely improbable that there are two such kinds; and if they are, we must abandon for ever the hope of having a single scale to represent all the facts of loudness; we shall need at least two independent scales. Nobody who makes assumption *C* must criticize the decibel scale on the ground mentioned on p. 161.

Assumption *C* does not seem to have been adopted in quite the simple form in which we have stated it, but it is involved in the procedure of Ham and Parkinson† and of Churcher, Davies and King‡. They ask observers to pick out sounds which

* One of us has been guilty of it himself; see N. R. Campbell, *Proc. phys. Soc.* **45**, 590 (1933). The mistake is made easier by Fechner's errors, discussed on pp. 163–165.

† L. B. Ham and J. S. Parkinson, *J. acoust. Soc. Amer.* **3**, 511 (1932).

‡ *Loc. cit.*

ppear to them to stand in some prescribed ratio of loudness, such as $\frac{1}{2}$, $\frac{1}{3}$, $\frac{1}{5}$. They hen use (11) combined with (14), assigning the prescribed value to a. It is to be bserved that this is equivalent to using (12) in the form $F\{S_0(x)\} = \rho . F\{x\}$, where is the prescribed ratio.

This procedure, though it might be well adapted to the establishment of a purely onventional scale, is ill adapted to discover facts. For the question which is put o the observer really consists of two parts: (i) Can you pick out pairs of sounds vhich appear to you to be equally related? (ii) If so, and if such pairs are to be epresented by numerals in a constant ratio, what ratio do you prefer? If a definite atio is named, the question substituted for (ii) is "leading," especially if fact (j) s true. For if there is a relation more definite than any other, an observer asked o pick out sounds "in a 2 : 1 ratio" will tend to select this relation, and will have o call it 2 : 1; no other alternative is offered him. The correct procedure to elucidate acts would be to separate the two parts (i) and (ii); it is then by no means certain that observers who all picked out the same pairs of sounds would all name exactly the same ratio; indeed experiments on mental estimates suggest strongly that they would not.

Even if they did, it would still be possible, as suggested on p. 158, that they were all using the same analogy and that this analogy might be false. Thus they might all pick out 2 : 1, because it is the "simplest" ratio between finite numerals and therefore appropriate to represent a simple relation in loudness. But the simplicity of 2 : 1 is based on the facts of counting, the determination of number, an additive magnitude having certain unique properties. The facts on which counting is based have no analogues in the field of loudness; an analogy based on simplicity is false. Again when they say that sound B is to sound A as 2 : 1, they may be judging that the relation between B and A is the same as that between A and some sound (e.g. silence) to which they instinctively attribute the numeral 0. But if that is so, they are simply blundering when they assign to C, which bears to B the relation of B to A, the numeral 4; they ought to assign the numeral 3. Indeed it is highly suspicious that the relation of 2 : 1, which plays so large a part in these investigations, is the most ambiguous that could be selected; no other leaves it so doubtful what numerals are to represent the higher powers, in the logical sense, of the relation.

We must not be misunderstood. We do not deny that by the procedure discussed here it is possible to obtain a scale of loudness which satisfies completely certain people and does not greatly offend anyone. What we do deny is that sufficient evidence has been advanced to prove that this concordance is derived from facts which must have universal validity, and not merely from generally accepted conventions and customs, the validity of which might be challenged by others with different training and habits. The importance of this distinction is discussed on p. 168.

(iii) *Just-perceptible differences*

We next turn to an entirely different procedure, now generally discredited, but historically important and still exerting a baneful influence. We shall assume the

Fact (k) following fact (k). Given any sound X of intensity x, another sound Y can be found of greater intensity y, such that Y is louder than X, but any sound Y' of intensity between y and x has the same loudness as X. X and Y then stand in the just-perceptible-difference relation. We shall assume further that y is a single-valued and monotonic relation of x within a wide range of conditions, subject only to a determinate experimental error. This second assumption is by no means universally granted; some hold that y is affected profoundly by insignificant changes in experimental conditions, such as fatigue from previous observations. But, since the just-perceptible-difference method is clearly worthless unless the assumption is true, we shall make it for the sake of argument*.

One way to use this fact would be to plot y against x and use the resulting curve in precisely the same way as one of the curves of family S in the equal-relation method†. The resulting method would be affected by the arbitrariness that was discussed on pp. 161, 163. But that is not the usual procedure. The usual procedure,
δx due to Weber, is to plot $(y-x)$, which is usually called δx, against x or the mean of x and y. Weber suggested that the resulting curve was a straight line, so that

$$\delta x/x=b \qquad \text{......(15)},$$

b where b is a constant. It is now agreed that his suggestion is false and that the curve is of some other form

$$\delta x=W(x) \qquad \text{......(16)},$$

W where W is a function of which the algebraic form cannot be defined at present; but, if the facts are facts, (16) is a perfectly legitimate way of representing them.

But the next step, taken by Fechner, is not legitimate. Fechner replaces b in
δX (15) by δX, a constant difference in a magnitude X, loudness, which *ex hypothesi* he has not yet measured. He then writes $\delta X/\delta x=1/x$, identifies $\delta X/\delta x$ with a derivative, applies to the equation the operator $\int_0^X dx$, and arrives at the equation

$$X=\log x-\log x_0 \qquad \text{......(17)},$$

giving a relation between X and x of the form $X=F(x)$ which we are seeking. His followers have applied the same procedure to (16), performing the integration graphically and arriving at

$$F\equiv\int_0^x \frac{1}{W(x)}\,dx \qquad \text{......(18)}.$$

It is as difficult to explain precisely why this is wrong as to explain why it is illegitimate to translate a French book with a German dictionary. The operator

* The question of fact, namely whether just-perceptible differences are unique, has become hopelessly entangled with the question of interpretation, namely whether Fechner's law expresses the fact. Those who deny the fact maintain, as a second line of defence, that Fechner's law does not express it; and those who see no fault in Fechner's argument are disinclined to admit that the facts he assumed are not facts. Here we want to separate the two questions completely.

† Indeed the curve might be the extreme member of that family; it would be so if all just-perceptible differences were recognized by the ear as similar intervals. It is doubtful whether they are so recognized. If the just-perceptible-difference curve were the extreme member of the family it would provide a valuable test of the legitimacy of assumption A; but, as its membership may be doubted, the test is not stringent.

$\int_0^x dx$ is applicable only if the argument is a derivative, the limit of a ratio between two variables when the denominator tends to zero. But $\delta X/\delta x$ is simply not a derivative; δx is not a variable which can tend to the limit zero—it is always finite; and δX is, according to Fechner's assumption, not a variable at all but a constant. The whole procedure is mathematically absurd.

It is equally absurd physically; for even if an equation is mathematically integrable it does not follow that the equation resulting from its integration will be significant physically. An equation $y=f(x)$ states a numerical law only if the symbol $=$ represents not only a numerical relation between the values of the magnitudes y, x, but also a physical relation between the systems to which they relate; generally they are magnitudes of the same system. If the equation is transformed mathematically it will acquire a new physical meaning only if this physical relation is replaced by some other. Whatever the physical relation, we can perform the mathematical operation $\int_0^x dx$ and arrive at a new equation

$$U=\theta(x) \qquad \text{......(19),}$$

where $\theta \equiv \int_0^x f(x)\,dx$; but it is not necessary that U should be a magnitude and (19) a numerical law. Thus if y is viscosity and x temperature, the integration will be physically meaningless; U will not be a magnitude and (19) not a law.

Indeed there are only two conditions in which integration is, in this sense, physically significant. The first arises when y is a derivative derived by differentiating (19); then integration will reverse the process and give the original equation. Thus y may be a coefficient of expansion with temperature and x temperature; (19) then gives the relation between the whole expansion and the change of temperature. But here y must have been, in effect, measured by using this relation. The second condition arises when additive quantities are concerned. Thus if y is the cross-sectional area of a tube and x the distance from one end, (19) gives the relation between x and U, the total volume from that end to x. But here $y.dx$ must be (or rather be proportional to) an additive magnitude, namely volume.

Neither of these conditions is present in the Fechner integration. dX/dx is certainly not a derivative obtained by differentiating some relation between X and x; for the argument starts from the hypothesis that none is known. And X is not an additive magnitude; for the existence of just-perceptible differences is inconsistent with additiveness; compare fact (*d*).

The preceding sentence is important, for an attempt has been made to avoid Fechner's mathematical absurdities by representing the just-perceptible-difference scale as one obtained by adding a number of equal steps from the threshold. But measurement by a number of equal steps is valid only if the magnitude is additive; a length can be measured by a number of equal steps, a density cannot.

The truth is that Fechner was simply wrong; his blunders have survived merely because text-book-writers copy one another*. The relation $X=F(x)$ obtained by

* Similar criticism of Fechner's work has been made before. See, for instance, F. C. Bartlett, *Discussion on Audition*, p. 128 (Phys. Soc. 1931).

his method is at best a very obscure way of representing the relation between δx and x, both of which are physical magnitudes and neither of which is conveniently identified with loudness.

(iv) *The thresholds*

The best-known scale of loudness has been left until last because the facts that it expresses are much more limited than those on which the other methods depend. These other scales try to express relations of which any sound may be one term; the decibels-above-threshold method does not express any audible relation between two sounds; it expresses only a single fact about one single sound of each pitch, namely that there is a sound such that any sound of less intensity is inaudible. The assignment of numerals to represent other sounds depends wholly on physical relations, namely ratios of intensities, not on relations in which the ear is necessarily concerned (see p. 169). Such relations, involving all sounds, might have entered if it had been found that sounds of different pitch and the same distance-above-threshold were equally loud; but that turns out not to be true. Accordingly the decibel-above-threshold method expresses fewer facts than other methods and is to that extent more arbitrary. In its latest form, where the threshold is chosen arbitrarily and not as a result of experiment, it expresses no facts at all except those common to all these methods, namely that, among sounds of the same quality, loudness is a single-valued and monotonic function of intensity.

§ 3. SOUNDS OF DIFFERENT QUALITY

In this field the best that we can hope to achieve is a scale determined by facts valid for some "normal" group of observers; it would be impossible to obtain a scale applicable to frequencies as high as 10,000 equally valid for old men and children. But since this difficulty is overlain by others there will be no need to insist on it.

If we had concluded that some method determined uniquely a function F_p for a single tone, we should have had to inquire now whether, when p varied, the variation of F_p could be represented by the change of a single parameter; further, we should have had to inquire whether the loudness of complex sounds could be represented by some general function F of equation (3), which reduces to F_p when all components but one vanish. But since we have failed in our search for F_p we must proceed, if we can proceed at all, in the reverse direction. We must inquire whether the study of tones of different pitch and of complex sounds help us in our search for F_p.

It is possible that it might. Consider again the equal-relation method. It depends on assumption A, that there is a function $S(a, x)$ involving a single parameter for sounds of the same pitch. If that assumption were fully established for all pitches, it would be natural to suppose that this parameter might be a function of pitch p, so that all the families S for all pitches could be represented by a single formula

P

$$y = S\{P(p), x\} \qquad \text{......(20)},$$

where P must be single-valued but need not be monotonic.

Assumption

It would not be as easy to produce indirect arguments for this assumption, which may be called D, as for assumption A; it seems to require that an algebraic form for $S(a, x)$ should be found. But if it could be established, interesting consequences might follow. From (11), on substitution of $P(p)$ for a, pairs of functions ϕ and F might be formed as before; but since the conditions to be fulfilled (if assumption D had already been established) would be more stringent, the number of permissible pairs would be reduced. It *might* turn out that only one pair was permissible; then $X = F(x)$ would be a true measure of loudness, free from arbitrariness. Whether it would turn out so is a question of fact, which it is impossible to investigate until assumption D has been established; but this appears to be the direction in which it seems most profitable to seek a further advance by means of the equal-relation method.

What may seem at first sight to be an alternative method is that of Fletcher and Munson*. The problem with which they are concerned primarily is this. Given all the (p_k, x_k)'s for a sound, find by calculation the intensity x_r of a reference sound of fixed frequency p_r which is equally loud. Any solution of that problem—and we assume that Fletcher and Munson have solved it—requires the finding of an F which satisfies the test given on p. 155 and has the same value for different sets of (p_k, x_k) which correspond to equally loud sounds of different quality. But the solution does not require that the F found should satisfy the tests given on p. 156.

As a matter of fact, Fletcher and Munson's solution does not satisfy the test that F reduces to F_p for a pure tone. If this were all, perhaps the conclusion should be that this test is too strict and that "reduction" in the mathematical sense is not necessary; all that is required is that a set of functions F should be set up covering between them the whole field of pure and complex tones. We are not entirely ready to admit that conclusion; it seems to us that, if the scale is to be satisfactory, the loudness of a pure tone of intensity x ought to be calculable from the loudness of a pure tone of the same frequency and intensity 1; or, in other words, that the tone of intensity x ought to be capable of being regarded as a complex of x pure tones of intensity 1. For if that is not so, the scale does not express the very important fact that intensity (unlike loudness) is additive. It seems impossible that this condition should be fulfilled unless F reduces mathematically to F_p.

But that is not the only reason why Fletcher and Munson's scale does not solve the problem with which *we* are concerned. Their F_p is simply decibels above a given arbitrary threshold. If they had shown that it was impossible to arrive at any function F, passing the test given on p. 155, unless F_p had this form, they would have produced a fact affording a reason for adopting their scale F_p and therefore F. But it does not seem that they have shown this. So far as we can ascertain, they could have carried through precisely the same procedure if they had made F_p for the reference tone simply the intensity x_r. Of course they would have arrived at a different F; but they would have arrived at some F valid over the same range. If that is so, they have produced no reason such as we are seeking for preferring the scale F_p,

* *Loc. cit.*

which they adopt, to a scale in which, for pure tones, loudness is simply identified with intensity.

There is then no evidence that we can find in the study of complex sounds facts which will serve to select a single scale of loudness from the infinite number, all single-valued functions of each other, consistent with facts (*a*)–(*f*). We still have to admit that all scales contain an arbitrary element at present irremovable.

§ 4. GENERAL CONCLUSIONS

Are we then to conclude that the scales are all useless? This question must be relative to some purpose. One of the chief aims of pure physics is to arrive at theories to explain facts. In the course of its long history, it has always turned out that theories can be found only when the facts, if expressed numerically at all, either are expressed in terms of "true" scales, based completely on facts except for an arbitrary factor, or else are equally expressible on all scales alternative according to the principle laid down on p. 156. Why this should be so is a mystery, but there are abundant examples; for instance, there could be no adequate theory of heat before the absolute scale of temperature was developed.

If then our purpose were to develop a fundamental theory of audition, we might well despair of making an advance so long as arbitrary scales are necessary and no others are available. Or again, if our purpose were merely to describe the facts in a compendious form, and if the purpose could not be attained at all unless all the facts were described in a single compendious form, we might equally despair; for in physics the only laws that describe all the relevant facts are those which are found to have a theoretical basis. Empirical laws are always of limited scope.

But that is not the position. A scale of loudness is required by the engineer concerned to reduce the annoyance caused by his machinery; by the designer of radio-receiving sets concerned to obtain faithful reproduction of music; by the telephone engineer concerned with masking of speech by interference. The first is concerned mainly with the upper range of audibility and with complex sounds, the second with the lower range and with nearly pure tones, and the third with the lower range and complex sounds. Their fields are so distinct that they very seldom need to correlate their observations. There is no reason why they should employ the same scale. While each believes that the scale that suits him is a "true" scale, imposed by ineluctable facts, they will never agree to differ; each will try to convert the other to what he believes to be the one and only truth. But if they would realize that all scales are arbitrary and that one is just as little true as the other, toleration would become easier. Whether any of the scales that have been proposed is actually well adapted to any of the various purposes for which scales are required, and, if so, which scale is adapted to which purpose—these are questions on which we express no opinion. But we cannot help believing that if the latitude which the presence of an arbitrary element permits were generally recognized, progress towards the adoption of standard scales, each for its own particular purpose, would be more rapid.

There is a final question that we are loth to attempt, but is almost sure to be

raised in any discussion of this matter. Is there any reason why it should be impossible to find a true scale of loudness?

The only way to answer this is to examine the differences between physical magnitudes which can be measured and psychological magnitudes which, we feel, ought to be measurable, but which actually are not. The difference that appears to us most distinctive is this. Physical magnitudes are measured by processes depending on sensory judgments which can be made with any sense organ. Sight is usually employed, but that is merely for convenience. Given suitable apparatus a blind man could determine for himself a weight, a density, or even magnitudes closely connected with vision, such as wave-length or intensity of radiation. But it is inconceivable that he could measure colour or brightness. These terms depend for their meaning on visual judgments; anything that a blind man could measure without consultation with a sighted man would, *ipso facto*, not be colour or brightness. Psychological magnitudes are those which, if they are to be measured at all, must involve judgments made with a single sense organ. Loudness is one of them, for its meaning is derived from fact (*a*).

An equivalent but cruder way of expressing the difference is to say that the physical magnitudes are properties of all matter, while the psychological magnitudes are properties of the particular pieces of matter that form our sense organs. If that is so, some explanation appears of the fact that psychological magnitudes do not possess the properties that would make them measurable. They are not additive because our sense organs are not constituted in such a way as to make them so; additive mechanisms are generally very simple and our sense organs are very complicated. Again they do not appear, generally, as constants in a numerical law between physical magnitudes, because the only laws in which they could appear as constants would be those involving the operation of our sense organs, and the physical magnitudes characteristic of these organs are not open to observation in the living and conscious subject.

Further we can see why some psychological magnitudes such as musical interval, discussed on pp. 158, 159, are measurable as *B* magnitudes. The reason is that they do not involve this structure so intimately as the others. They are all of the nature of differences; the difference between the effects of two stimuli on a mechanism may be more nearly independent of the structure of that mechanism than the effects themselves; it may be more nearly a property of matter.

Does this mean that, if we knew more about the operation of our auditory organs, we could hope to measure loudness on a unique scale? Our answer is negative, so long as the general outlook on the problem remains what it is at present. Loudness can never be measured as an *A* magnitude, because it is not additive. That is a fact which nothing can change; it would not be changed if we found that loudness was intimately connected with some *A* magnitude characteristic of the auditory organs, such as the rate at which impulses pass along some nerve system. The most that any further knowledge of these organs could do would be to correlate loudness with some magnitude *H* characteristic of these organs, so that the order of loudness would be the order of *H*. If *H* were a constant in a numerical law of these

organs, loudness could then be measured in the scale of H as a B magnitude. But it does not follow that this scale would be satisfactory; it might be unsatisfactory just as the intensity scale is unsatisfactory for pure tones, although for them intensity is correlated exactly with loudness. If it were already satisfactory, we should have made no advance; if it were unsatisfactory, it would not become satisfactory because we had discovered that H and loudness were correlated.

This is the crux. At present the reasons that determine whether a scale is satisfactory are based on facts (g), (h), (j). If these do not determine the matter, doubt can be removed either by finding other facts of the same kind, or by ceasing to attribute importance to these facts and being as indifferent to scales as we are in respect to some physical magnitudes. There is no third choice. It is perhaps too early to despair of the first; but in case it proves ultimately impracticable, it is worth while to point out that the second is open. If we could only rid ourselves of the feeling that loudness ought to be additive and that equally related sounds ought to be represented by numerals in some simple arithmetical relation, disputes concerning scales of loudness would vanish; the only remaining problem would be to measure the loudness of complex sounds on some scale, no matter what. We should concentrate on F and cease to be interested in F_p, which might be any single-valued function of intensity. Then it would be quite possible that the best way to find F would be by means of H as intermediary. Here we find ourselves in complete agreement with what we understand to be the view of Fletcher and Munson, even if we cannot accept all the arguments by which they support it.

§ 5. ACKNOWLEDGMENT

We desire to acknowledge the help we have received from discussions with our colleague, Mr D. A. Oliver.

DISCUSSION

Dr J. H. Shaxby. Two papers in this month's *Proceedings of the Royal Society* B describe experimental investigations made recently by F. H. Gage, in this institute*, on the possibility of forming scales of sensation. The sensations chosen were brightness and loudness, and the conclusion reached in both cases is that the method of dividing a given sensation-interval into two equal parts does not lead to a consistent scale. Briefly, if we thus bisect the whole interval and then each of its halves and finally similarly divide the interval between the $\frac{1}{4}$-way and $\frac{3}{4}$-way points, we do not arrive again at the same stimulus as that for the original $\frac{1}{2}$-way point. The term bisection is not to be taken as implying equality of differences of loudness or brightness, but only equality of the two "intervals" in the sense in which Campbell and Marris use that term. This implies the failure of other methods of founding scales, such as the just-perceptible-difference step method, since the number of steps to bridge the interval to a brightness (for instance) lying half-way between two

* The Physiology Institute, University College, Cardiff.

given brightnesses differs according to the means adopted for defining the stimulus giving the bisecting brightness, and this is indeterminate.

A fact that emerged in the course of the work is of great interest. Many observers have in different researches halved sensations, as for instance in Churcher's paper referred to by the authors and in older work on brightnesses. We tried a somewhat different procedure, namely the bisection of the interval between a moderate brightness and a very feeble one, repeating the tests with a progressively weaker dim light until the threshold was nearly reached. The limit of this, the bisection between the fixed brightness and zero brightness, would, one might have supposed, have been equivalent to halving the fixed brightness. On the contrary, the bisecting stimulus was but little above the feeble stimulus, and seemed to approach it as the latter approached the threshold. It is clear that the judgment made in this way is quite different from that in the halving experiments in which the influence of the comparison weak light is lacking.

The observation is further evidence against the relation $\phi=v-u$, and possibly, though not certainly, in favour of $\phi=v/u$.

While all who tried Gage's bisections agreed that the judgment was definitely possible and comparatively easy to make, none felt any certainty as to what they were doing. Whether they were obtaining equal differences or equal ratios of loudness or brightness they did not know: all they were sure of was that there was some definite equality involved. In the authors' phrase, the two intervals were equal. My own impression was that it was *not* the differences which were equal, but that the difference between the greatest and medium sensations was larger than that between the medium and smallest, although the intervals were equal.

The authors are not the first to demolish the mathematics of the Weber-Fechner law, but they have done it with unusual thoroughness. They do not add the comment, itself not new, that an equation expressing the relation of sensation to stimulus as a continuous function and derived from another equation formally asserting the discontinuity of that relation, is indeed remarkable.

But despite the mathematical illegitimacy of its origin the law does express, what the decibel scale of course also expresses, the fact that sensations seem to run more or less parallel to the logarithms of the energies of the stimuli which evoke them, while commonsense immediately rejects the suggestion that their values are proportional to the energies themselves. This surely is the fact, revealed by one scale and concealed by the other, which has given such undeserved prominence to the decibel scale and misled so many into regarding it as a scale of sensation and not merely as one of stimulus.

Dr J. E. R. Constable. One small point may perhaps be added to the valuable analysis which the authors have given. I refer to their suggestion that the loudness-doubling upon which other authors' work is based is probably only a "most-easily-recognized interval". It has been argued that the agreement between loudness-doubling results and the monaural-binaural measurement is an argument in support of the possibility of recognizing a factor of 2 in loudness. The experiments un-

doubtedly show that a sound when heard by two ears is perceived as louder than when heard by one. It does not appear, however, to have been demonstrated that the loudness is, in fact, twice as great in the former as in the latter case. Might it not be possible that the agreement between the experiments is due to the fact that the difference between monaural and binaural listening is a relation with which everyone must be familiar and might merely constitute the most-easily-recognized interval to which the authors refer?

As regards the practical aspect of the authors' results it may be remarked that even if the supposed loudness-doubling is only a most-easily-recognized interval, the usefulness of loudness scales such as that put forward by Churcher, King and Davies would not necessarily be impaired; for, from the practical point of view, a scale made up of a series of most-easily-recognized intervals is as significant as one based upon a succession of loudness-doublings.

Mr C. R. Cosens. An alternative to the procedure outlined in the second paragraph on page 163 suggests itself, which would appear to be free from the leading-question objection. Let there be two sounds, A and B, whose loudnesses are a and b respectively. The observer is required to determine what, in his opinion, is the value of the ratio a/b. Provide two adjustable sources of sound X and Y (say loud-speakers fed from potentiometers) with an arbitrary calibration of their volume-controls (say the angle θ through which the control-knob is turned).

The observer first adjusts the knob of X until it appears to give the same loudness as A, and $\theta(a)$ is recorded. He is then asked to adjust Y until it appears twice as loud as X, and $\theta(2a)$ is recorded. X is then adjusted to be twice as loud as Y, giving $\theta(4a)$. This doubling process is repeated until the sound is undoubtedly louder than B, and a graph can then be plotted connecting loudness as estimated with the arbitrary variable θ.

Sound B is then introduced, for the first time, to the observer, who is asked to adjust one of the sources, X or Y, until it is of the same loudness as B, and $\theta(b)$ is recorded. The value of the loudness of B in terms of multiples of a can then be interpolated from the graph.

A few rough experiments were made on these lines some years ago, the tuning-note of the B.B.C. and a wireless receiver being used.

θ was the reading of a square-law voltmeter connected across the loud-speaker windings, the volume was adjusted to a definite θ in each case, and observers were asked to turn up the volume-control until the sound appeared to them to be twice as loud. Results of course varied considerably, but a more or less linear relation with θ was shown, leading to the conclusion that "twice as loud" meant four times the voltage across the loud-speaker, i.e. 16 times the power expended, so that for this particular pure note and starting from one particular loudness, the mean observer considered the loudness to vary as the fourth root of the power converted into sound.

Dr R. T. Beatty. During the year which has elapsed since Dr Campbell read a paper on the measurement of sensation before this Society he has made a psychological advance, for the operations which he then described as "guessing" he now

romotes to the rank of "mental estimates," a term which is less suggestive of a ame of pure chance.

The available experimental evidence as to subjective loudness may be sumarized in a single sentence as follows. *In the domain of sensations of loudness the rain can perform the operations of multiplication and division.*

Four investigations* are to hand in which observers were asked to alter the loudess of a sound to some multiple or sub-multiple of this loudness. For this operaion they depended solely on their own subjective intuitive judgment as to what the nagnitude of the change in loudness might be. When the four curves are compared he maximum deviation from the mean curve does not exceed 8 decibels in a range of 80 decibels, i.e. an intensity-ratio of 6 : 1 in an intensity-range of a hundred nillion to one. If this is guessing, the guesses are remarkably consistent. They are urther supported by a deduction from the work of Fletcher and Munson†, a deduction which Dr Campbell has failed to make, but which has now been pointed out by Mr Churcher. A sound at a level of 100 decibels applied to one ear seems of the same loudness as a 90-decibel sound applied to both ears. On the assumption (derived from consideration of the numbers of nerve impulses) that binaural hearing gives twice the loudness of monaural hearing, we thus find that a reduction of 10 decibels gives a loudness-ratio of 2 : 1, and so a (stimulus, sensation) curve can be plotted. The agreement of this curve with the curve deduced from mental estimates is remarkable. The ratio 2 : 1 has now hurled itself so violently into the evidence that it is difficult to ignore it.

But the authors simply cannot understand how the idea of number can enter into sensations of loudness. I had hoped that my remarks‡ on the 1933 paper might furnish a clue. It is unnecessary to repeat the suggestion then made except to say that it indicated that the correlation between eye and hand involved in handling weights, together with the linear relation between weight and afferent nerve impulses, furnish a numerical relationship which might be transferred to the auditory or visual receptor system.

Several attempts have been made to prove that Fechner's law is mathematically unsound, but if our sensations were additive Fechner's law would be valid, and indeed the analysis given in this paper can by a slight modification in the wording be used to prove that it *is* valid. This paradox is due to the fact that Fechner's integration is not a pure mathematical construction. Its history serves to warn us that the algebra of sensation is not necessarily the same as that of physical magnitudes.

It is difficult to discover what, apart from conclusions of despair, the authors' general conclusions are. One remark however, to which they seem to attach some importance, deserves comment. They say, "If we could only rid ourselves of the feeling...that equally related sounds ought to be represented by numerals in some simple arithmetical relation, disputes concerning scales of loudness would vanish."

* Richardson and Ross, *J. gen. Psychol.* p. 288 (1930). Ham and Parkinson, *J. acoust. Soc. Amer.* **3**, 511 (1932). Geiger and Firestone, *J. acoust. Soc. Amer.* **5**, 25 (1933). Churcher, King and Davies, *J. Instn elect. Engrs* (October, 1934).

† Fletcher and Munson, *J. acoust. Soc. Amer.* **5**, 82 (1933).

‡ Beatty, *Proc. phys. Soc.* **45**, 572 (1933).

Why should we try to rid ourselves of this feeling, since five investigations* are on record to show that such numerical relationships actually exist?

Mr W. F. Floyd. I cannot agree with the desire of the authors, expressed by one of them (G.C.M.) in his introductory remarks, to place this discussion outside the field of psychology. Loudness as defined in the paper in (*a*), or as defined in any admissible way, is essentially a mental estimate. The estimate may be a psychological correlate of the physiological change resulting from the stimulation of the outer ear, or it may be a simply definable function of the endo-somatic response dependent on whether we reject or accept a purely behaviourist (physico-chemical) description of human effort. Mental estimates are acceptable, statistically, to psychologists and form reliable measures, provided that the usual statistical conditions are satisfied, namely that the sample is large and random and the experimental methods do not arbitrarily bias the results. If the authors believe that human effort may be explained in terms of physics and chemistry only, their problem is proper to the physiologist and may be solved by experiments in which a measure is made of the potential-fields and barriers in the nervous system and the cortex†. If they believe in mind they must accept mental estimates and the statistical methods of the psychologist, and use the mean value of mass data, with the variance, to set up a loudness scale and so determine their function (3).

The difficulties which the authors conceive in the construction of a scale for pure tones of a single frequency by the method they designate "mental estimates" are difficulties of experimental technique only, which occur in all psychological experiments. It appears to me, as I interpret the authors' statements, that the method § 2, (i) which they would discard is based directly upon their first fact of audition, (*a*). Indeed, when the experimental technique is correct, the method is a restatement of the last sentence of (*a*).

In place of equation (16) of the paper I would suggest the relation

$$\Delta X = \phi\,(x, \Delta x) \qquad \text{......(16}a\text{)}.$$

This does represent the facts, subject to the condition that ϕ is monotonic and single-valued. To every variate of ΔX there correspond unique variates of the variables x and Δx, and we have thereby avoided the error of assuming that all the ΔX's are equal, an error which the authors condemn but which they commit on p. 165 when they say "The preceding sentence is important, for an attempt has been made to avoid Fechner's mathematical absurdities by representing the just-perceptible-difference scale as one obtained by adding a number of *equal* steps from the threshold."

The term "equal" is applicable where two stimuli, occurring alternately, arouse physiological changes which have undistinguishable psychological correlates. The term must be warily applied to increments in perceptions. A few experiments in this direction have shown the exponential relation between stimulus and percept to be approximately true for the subjects examined.

* See footnotes *, †, on p. 173.

† *Vide* the work of Adrian, Lashley, and others.

Mr B. G. Churcher. In the introductory part of the paper the authors have performed a useful service in laying down with precision certain fundamental propositions, with which most acoustical technicians will be, I imagine, in complete agreement. The realization that loudness is not additive eliminates the necessity for considering two scales, viz. the decibel loudness scale and the scale based on the sum of the observed just-perceptible increments of intensity. As the authors point out, it is theoretically possible to use many types of scale. But, from both scientific and utilitarian points of view, it is desirable to use a scale in which the numerical value assigned to a sound is in direct proportion to the loudness-sensation evoked by that sound. If such a scale is possible, then to use any other kind is to create unnecessary complication. To obtain such a linear scale has been the object of my colleagues and myself. Since most noise-measurements are stated in terms of the intensity of the equally loud reference tone above threshold intensity, the relation between decibels above threshold and loudness for a pure 1000-cycle tone is of particular interest. Since loudness is a sensation, the fundamental criterion of relative loudness is a representative individual's assessment of it. Hence the method of mental estimates of relative loudness is an important guide even if it is not the most precise method of arriving at the (decibel, loudness) relation. Loudness scales which are in conflict with the data of representative mental estimates are open to suspicion. The authors have apparently misunderstood the method of mental estimates employed by my colleagues and myself and I do not think that their objections hold. If the five investigations that have been carried out on this subject* be compared, it will be found that, one being omitted, a representative curve may be drawn relating decibels and loudness on a scale, such that the loudness evoked by 100 db. above threshold is designated 100 and that any individual investigation deviates from the curve by not more than 8 db. (A lantern slide was shown.) An explanation of the faculty of making intuitive estimates of relative loudness is afforded by the nerve-impulse theory of audition†, which also explains the phenomena observed when a sound of given intensity is applied to one ear and alternatively to both ears. The theory indicates that the loudness is doubled in the latter case. If the relative intensities for monaural and binaural listening for equal loudness are measured, a (decibel, loudness) relation can be deduced. Data from two independent sources show good agreement and a representative curve can be drawn. (A slide was shown.) Fletcher and Munson‡ have proposed another method based on their theory of the loudness of complex sounds. But the general validity of their theory is doubtful as no distinction has been made between sounds with harmonically related components and those with non-harmonically related components. The enhanced loudness obtained with harmonic relations is beyond doubt and has been

* L. F. Richardson and J. S. Ross, *J. gen. Psychol.* 3, 288 (1930). D. A. Laird, E. Taylor and H. H. Wille, *J. acoust. Soc. Amer.* 3, 393 (1932). L. B. Ham and J. S. Parkinson, *J. acoust. Soc. Amer.* 3, 511 (1932). P. H. Geiger and F. A. Firestone, *J. acoust. Soc. Amer.* 5, 25 (1933). B. G. Churcher, A. J. King and H. Davies, "The measurement of noise with special reference to engineering noise problems," *J. Instn elect. Engrs* (October, 1934).

† Dr R. T. Beatty's remarks: "The measurement of visual sensation," Discussion, *Proc. phys. Soc.* 45, No. 249, p. 565 (July, 1933).

‡ Harvey Fletcher and W. A. Munson, *J. acoust. Soc. Amer.* 5, 82 (1933).

recently confirmed by Fletcher himself. Tone-combination methods are not, therefore, admissible for the construction of a loudness scale. If now the representative mental estimates and monaural-binaural curves be compared, close agreement over a loudness-range of 200 down to 0·4 will be found. (A slide was shown.) The final curve based on these results is quite well represented by the law $L = d^5 \times 10^{-8}$, where L is the loudness and d the decibels above threshold. (A slide was shown.) The concordance of the nerve-impulse hypothesis, the mental estimates and the monaural-binaural results seems to indicate that a representative, if not a unique, loudness scale can be formulated.

Mr W. West. Experience with subjective tests (i.e. tests of what the authors term "psychological magnitudes") has led me to regard these as comparative only, requiring skill to ensure that the comparison is both apt and fair. I am therefore surprised to learn (p. 168) that telephone engineers and others believe loudness scales to be "true" scales. The most universally employed loudness scale is, as I regard it, no more than a means of ascribing numerical values to one aspect of noisiness. It has been given the name "equivalent loudness" in order to avoid confusion with the general use of the word "loudness." It equates, by subjective test (comparison), the loudness of a noise with that of a standard tone (arbitrarily selected as 1000 c./sec.), and states the equivalent loudness of the noise in terms of the sound pressure of the standard tone (an exact physical measurement). It seems to me that this method is referred to, but misrepresented, in paragraph 2, (iv) on p. 166. Details of standardization of this scheme may not yet be complete, but the principle is very generally accepted by engineers who have to measure noises. I would refer the authors to some tests and theories of loudness by U. Steudel*.

Mr A. J. Aldridge. The measurement of loudness is becoming increasingly important and obviously some scale is necessary. The authors have investigated various suggestions, and, so far as I have followed them, have indicated that no present method is satisfactory. They appear to doubt whether any true scale is possible. If this be so, the only thing which engineers can do is to carry on with the more or less arbitrary scales that have been selected as most suitable for the particular purposes they have in hand. On the other hand the authors, after having made a detailed examination of existing scales, and incidentally having pulled them to pieces, should be in a good position to suggest some improvement. I think we are entitled to ask them to do so.

Mr J. Guild. The authors have given a very clear treatment of the problem of loudness-measurement, as that problem appears to be envisaged by the majority of acoustical engineers. There are some points in the development of their arguments with which I do not completely agree, but as none of these leads to any material difference in the significant conclusions I will abstain from discussing them.

A somewhat startling conclusion which has emerged from the discussion which has followed the paper is that workers on acoustical problems, in so far as they were

* *Hochfrequenztech. u. Elektroakust.* **41**, 116 (1933).

presented by the participants in the discussion, hold views which are essentially fferent from those of Dr Campbell on the kind of association between number and agnitude which constitutes measurement. If the theory of measurement depended any way on a specialized knowledge of each field of phenomena in which it is pplied, one could hardly escape the conclusion that with this weight of specialist uthority against him Dr Campbell would be well advised to revise his ideas; but e principles which govern the association of numbers with magnitudes in order at the former may be *measures* of the latter and not merely more or less arbitrary umerical tickets, do not depend on the nature or properties of the thing measured. hey depend simply on the nature of measurement, and it is difficult to believe that ny physicist does not accept as fundamental the general principles on which the rguments in the present paper are based.

Mr L. V. K. Rein. It seems advisable to record an aspect of the work that has eached, to all intents and purposes, international acceptance. The particular field s the measurement of hearing-loss on the sensation-unit, decibel scale, as laid down or use in connection with the 2-A and 2-B (Bell Laboratory) audiometers. Over a teady series of researches carried out in this country and America during the last welve years, by otological and research groups, it has been definitely established hat not only may an observer's hearing-capacity be charted, but such progress has een made with the principle that libraries have been laid down where tens of housands of hearing-test charts are now kept, and characteristic audiograms for se in diagnosis have been conclusively arrived at as a result of these investigations.

The concrete application of the principle has been further established as a result of the recent two years' Hearing Aid Research Survey, carried out under the uspices of the Ministry of Health, whereby a series of instrument-specifications ased on performance curves calibrated in a state laboratory made possible the ractical application of the work. A series of these instrument-performance curves were converted to the sensation-unit scale, so that it became possible to match the ndicated hearing-loss curve with an instrument-performance curve compensating the patient's defect with an accuracy proved to be useful.*

As a result of careful comparisons, results have proved that a working co-operation with otologists using the half-intensity-unit (speed of decay of a tuning fork) method of testing is simple and satisfactory, and as a result of steady investigation where hundreds of cases have been tested by both fork and audiometer methods, provided that tests are carried out under silence-room or equivalent conditions, the discrepancies have been found to be practically negligible.

In the light of the vast field covered by the subject, an abstract taken from the conclusions arrived at during a similar discussion held recently by the German group in Berlin seems to express the present trend to reach a general agreement of international understanding in setting up an agreed unit and measurement, and whilst it is the writer's endeavour to take an unbiased view, the results of our practical principle are proving of such value that there is little likelihood of an immediate

* H. M. Wharry, *The Practitioner*, **129**, No. 5.

need for the re-establishment of a further unit in connection with our branch of the work, until an extensive series of investigations and experiments have been completed along the lines of those already in progress. These investigations seek to establish what has up to the present been a much neglected subject, the percentage of comparisons between an individual's hearing by air conduction in relation to his bone-conduction hearing-capacity.

A series of routine tests are being carried out at the present time in conjunction with three of our hospital research groups, and it is to be hoped that a new insight into an observer's ability to detect directional sounds may be established, even in the instance of a person with normal hearing, when a series of group categories based on the combined bone-conduction and air-conduction hearing curves have been made.

The object of bringing these facts to light is to eliminate any prospect of an erroneous idea being formed in connection with the measurements of hearing-capacity by established sound-intensities over a series of frequencies, and to offer the assurance that data are to-day available which prove beyond any question of doubt that the unit adopted in connection with our own field of work is adequately suitable, and offers a method of dealing in an entirely satisfactory manner with an individual's hearing-capacity, by the use of a measured scale which is now accepted and in use in centres and hospitals in the majority of cities throughout the world. The data, performance curves, and some thousands of audiograms taken in conjunction with otological and research groups, together with records of the work accomplished in America and Germany in this field, have been compiled and it is to be hoped that, when a little further investigation dealing with the aspect of bone-conduction hearing has been reached, full publication will be possible.

Mr H. Davies. In considering the method of mental estimates, § 2 (i), the authors have raised the question of the nature of the criterion involved in the decision that one sound is, say, twice as loud as another. It has been suggested by Dr R. T. Beatty* that for a pure tone the criterion is the number of afferent nerve impulses per second evoked by the stimulus and that we assign numerals to auditory sensations by analogy with the nerve impulse frequencies involved in the more familiar experience of lifting weights.

The idea is very suggestive and in a paper which is to be published within a few days† I have shown that if this hypothesis is accepted it is possible to construct a scale of loudness from the data of monaural and binaural minimum-perceptible changes of intensity by a method quite different from any of the four considered by the authors. Although it uses the minimum-perceptible-change data the method does not, of course, involve the process of integration of minimal changes to which the authors have administered so effective and so well merited a castigation.

In § 2 (ii) the authors have instanced a recent publication of Messrs Churcher and King and myself as indicating that there is a unique loudness relation, analogous to the octave, which is more readily identified than any other by the untrained ear. Of this I am not convinced. The work referred to is not entirely conclusive on this

* *Proc. phys. Soc.* **45**, 572 (1933).

† *Phil. Mag.* (Nov. 1934).

›int, since the only ratios employed were 2 : 1 and 4 : 1, and if there is any par-
:ular facility associated with the octave relation it will presumably also be in-
›lved in some degree in the double octave. Fact (j) therefore still remains open to
ıestion, but on the ground of introspective experience I doubt whether it is true,
:cept in so far as 2 : 1 is the simplest and most easily appreciated numerical ratio.
'he authors discuss the use of the 2 : 1 ratio and they suggest that its simplicity is
:levant to counting but that "the facts on which counting is based have no ana-
›gues in the field of loudness." But if, as has been suggested, the estimation of
›udness is essentially an estimate of nerve-impulse frequencies, then the process
analogous to counting and the simplicity of the ratio 2 : 1 has significance.

It will be agreed that in making an estimated partition of a straight line ruled on
aper the easiest division is into two equal parts and that the difficulty of the opera-
on increases with the complexity of the divisional fraction. Considering only the
ıental state of the subject making loudness-estimates, I rather think that the facility
onferred by particular ratios is much the same in the two cases. In fact, I should
xpect to obtain very similar results using estimates of $\frac{1}{2}$ and $\frac{1}{3}$ loudness, but would
xpect somewhat greater dispersions with estimates of, say, $\frac{3}{5}$ or $\frac{2}{7}$.

Whilst discussing this matter the authors refer to the case where a sound B is
› a sound A as 2 : 1 and a third sound C "bears to B the relation of B to A", and
hey state that to assign the numeral 4 to C is "simply blundering." As one who
ıas indeed done so, I am afraid that I am quite unrepentant. Surely this is only the
ıuestion of interval again. If the difference between C and B is the same as the
lifference between B and A, then, as the authors say, the numeral 3 should be
ıssigned to C. But if we accept

$$\phi(u, v) \equiv \frac{v}{u},$$

hen when we say that $C : B :: B : A$ we mean that the ratio of C to B is the same as
he ratio of B to A, in which case we must assign to C the number 4.

With regard to the facts set out in § 1 it has been usual, in the past, to consider
hat the loudness of a combination of sounds is independent of the relative phasing
›f the components. Recent work by Chapin and Firestone* however has cast doubt
ıpon this. If the effect of phasing is significant, then fact (e) as stated is not sufficient
ınd another variable must be introduced. Similarly relation (1) will not hold even
f A and A' are pure tones of the same frequency.

The authors have provided such a valuable addition to the literature of this
subject that it may seem ungracious to raise objection to minor points, but is not
the term "mental estimates" sufficiently general to cover also the methods which
the authors have classified as "equal relations"? I would like to suggest that the
expression "mental estimates" be reserved as a general description of all methods
involving estimates of sensations and that the particular process referred to in
§ 2 (i) be termed, say, "estimated interpolation."

Dr W. D. Wright. I believe that the problems which the authors are trying to

* E. K. Chapin and F. A. Firestone, *J. acoust. Soc. Amer.* 3, 173 (Jan. 1933).

investigate in this paper only exist because of two main misconceptions in their ideas. The first is that they do not yet appreciate the nature of a psychological entity. In the discussion on Dr Campbell's previous paper on the measurement of visual sensations, I had occasion to quote a statement of Mr Guild's and I feel that it is very necessary to repeat and emphasize it here. The statement was to the effect that "we cannot measure the magnitude of a sensation. We cannot experimentally isolate a unit of sensation from which to build up a quantitative scale of magnitude as required by all processes of measurement. This does not mean that we cannot make subjective estimates of the magnitude of a sensation; that in fact is the only means we have of obtaining quantitative information about sensations." If we wish to measure a length, we lay a calibrated scale of length alongside the unknown length and compare the two; if we wish to measure sensations, we must have a calibrated scale of sensation values before we can do so. Such a scale cannot be constructed and the attempt at measurement must be abandoned.

There is no evidence that this basic fact has been understood by the authors. One example may serve to illustrate this point. They contend that loudness is not additive because (p. 154) "ϵ can always be chosen so that if it is sounded simultaneously with a comparatively faint sound the combination is louder than that sound, but that if it is sounded simultaneously with a loud sound, the combination is no louder than that sound." This statement describes the effect, not of the addition of loudnesses, but of the addition of stimuli. The usual confusion between stimulus and sensation has occurred and the evidence proves nothing regarding the additiveness of loudness. The truth, as I see it, is that it is impossible to prove whether loudness is additive or not, because so long as we cannot isolate units of sensation it is impossible to carry out the experiments with the proper entities. I cannot myself conceive that it could be otherwise than additive, but that conviction is admittedly in no sense an experimental proof.

Although the authors entitle their paper "The measurement of loudness," they are not really concerned with such a measurement but are trying to find a comprehensive relation between stimulus magnitudes and sensation magnitudes, which is very different from the measurement of a sensation. But just as sensation-measurement is an impossibility owing to the nature of the quantity to be measured, so the derivation of a relation between stimulus and sensation is equally impossible, even if we could measure sensations, owing to the nature of the organ connecting the stimulus and the sensation. This is the second misconception of which the authors are victims. They seem to regard the sense organs as physical instruments in which, when you apply a certain stimulus at one end, you will always get the same response at the other. This, of course, is quite wrong. In the case of the eye, two differently coloured lights may be of equal brightness under one condition of the eye, but under another state of adaptation they may be of entirely different magnitudes. Evidently no stimulus-sensation relation can be postulated that will hold for both these conditions. It would be equivalent to using the same calibration curve for two very different instruments. Effects of this kind are possible for any of the sense organs, including the ear, and the facts must be faced. If they are faced, only one

nclusion is possible, namely that no stimulus-sensation relation can be found hich will be applicable to any but very limited variations of stimulation.

Dr L. F. Richardson. (i) What is a fact? It is customary in psychology to admit at phenomena peculiar to individual persons are facts. To demand universal lidity would be unscientific. Once this general principle is conceded, it follows at mental estimates of loudness are facts, and therefore respectable. And we turally pass on to consider the means and standard deviations of mental estimates ade by numerous persons*. (ii) In a recent paper F. H. Gage shows that bisection intervals of loudness leads to a remarkable misfit†. But Gage assumed that a xed stimulus always produced the same sensation, thus neglecting successive conast. Contrast may account for the misfit which Gage has found. (iii) Is it necessary blame Fechner so severely for using an integral as an approximation to a sum?

Authors' reply. (1) Two facts, unknown to us previously, emerged in the disussion. We will state and examine them in a way concordant with the rest of our aper. The first is that on which the monaural-binaural scale depends. The fact ere is that, given a sound B, another sound M can be found such that M heard onaurally is as loud as B heard binaurally. It is not a fact that to sound M must e attributed twice the loudness of sound B; that would be a fact only if binaural earing were the addition, in the physical sense, of two monaural hearings. But a ound heard by one ear decreases the power of the other to hear very faint sounds; here is therefore evidence of a decrease in sensitivity accompanying the first step, hich is fatal to additiveness. An analogy may help. If two similar bodies, each eated to 100° C., are dropped into water at 0° C., the rise of temperature is not wice that which occurs when one is dropped in. That is because the first body inreases the heat-capacity and therefore decreases the sensitivity. There is therefore definite reason for not calling the loudness of M twice that of B; if the monauralinaural scale agrees with the 2 : 1 scale, that is evidence against the latter.

The second fact is that of Mr F. H. Gage. What he has shown is that relations, lentified with equality because they are symmetrical, do not always (or usually) rove to be transitive. Equality in our discussion always implies transitiveness as vell as symmetry. (For this reason we do not accept the definition of equality ffered by Mr Floyd in his last paragraph.) The method of equal relations demands oth qualities, as we pointed out. Accordingly within the range of Mr Gage's facts hat method must be impossible; here we agree with him and Prof. Shaxby.

(2) For the rest we propose merely to underline and sometimes to expand statenents made in our paper, taking them in the order in which they stand there. In ddressing a Physical Society we felt justified in using "measurement" and its llied terms in the sense familiar to physicists, and ignoring entirely the very differnt sense in which they are used by psychologists. There seem to be two chief lifferences. First (if we understand Mr Floyd) psychological magnitudes are all vhat physicists call statistical magnitudes; the magnitudes we are concerned with

* L. F. Richardson and J. S. Ross, *J. gen. Psychol.* 3, 288–360 (1930).

† F. H. Gage, *Proc. roy. Soc.* B, 116, 103–119.

are not statistical, but simple. Secondly, if we understand Dr Wright, a "magnitude" can be "measured" psychologically, apart from all other magnitudes, whatever its qualities. In physics only additive magnitudes can be measured by themselves; all others are measured by means of relations between other magnitudes. Since loudness is not additive, it can only be measured in the second way. Accordingly establishing a relation between stimulus-magnitudes and sensation-magnitudes is not, for a physicist, a different thing from measuring loudness; it is the only possible way of measuring it.

As we said at the meeting, we do not believe that any part of our paper (except possibly § 4, which we do not defend seriously) has any relevance whatever to psychology. We address ourselves solely to physicists, who show by the discussion that they understand us except possibly in one particular: this is the significance of addition. The conditions stated in (1) are necessary, but not sufficient, criteria of additiveness. To state the sufficient criteria completely would take too long. Those who are inclined to doubt our statement that loudness is not additive should consider other cases of approximate additiveness, e.g. quantity of heat (mentioned above) or the output of a photoelectric rectifier cell.

(3) It is ungracious to quarrel with such a staunch supporter as Mr Guild; but we wish he would not speak of the association of *numbers* and magnitudes. The idea that numbers are associated with all magnitudes is the source of the fallacy which Dr Beatty upholds in his last paragraph. Only A magnitudes are associated with numbers; other magnitudes are associated only with *numerals*.

(4) We have never assumed responsibility for our "facts"; we have merely chosen those which, if they are facts, give the greatest chance of establishing a scale of loudness. If they are not facts, as some critics suggest, then the only conclusion is that the chance of establishing such a scale is even less than we estimated.

(5) In reply to a question put at the meeting, we define "intensity" by postulate; it is anything that makes (e) and (f) facts; both pressure and energy, measured in free space or in a telephone receiver, are intensities.

(6) We still refuse to define a fact formally. But it may help Dr L. F. Richardson if we point out that (in our sense) it is a fact that some people like blubber, but not a fact that blubber is delicious. Everyone agrees concerning the habits of the Eskimo, but not everyone shares their tastes.

(7) In discussing mental estimates we use "concordant" to mean "free from systematic divergence." In seeking agreement for technical purposes, it is permissible to ignore differences which, though systematic, are small. (Such differences are ignored, for example, in defining the colour-vision of a normal observer.) But if we are concerned to establish and use facts, small systematic differences are as bad as large ones. The facts that Dr Beatty and Mr Churcher themselves quote show systematic differences (a) between the estimates of different observers and (b) between the various scales which are said to agree. For some purposes the smallness of these differences may be important; for our purpose it is only their existence that matters.

Mr Constable's suggestion is interesting. The partial agreement between Mr

Churcher's various methods of obtaining a scale may arise from the use in all of them of the same arbitrary assumption concerning ϕ. If so, change of the assumption will alter all the scales in the same way and leave the agreement unchanged. For this reason even true agreement between several scales would not show that all are based on facts.

Our admission that systematic differences might be ignored for some purposes must not be taken as an admission that we regard those which Dr Beatty and Mr Churcher propose to ignore as actually negligible. On the contrary, in the branches of acoustics of which we have most experience, systematic discrepancies of 8 db. (a ratio of 6 : 1) would be very important indeed.

(8) We do not understand the relevance of Dr Beatty's theory of afferent impulses. If it is consistent with the facts that mental estimates differ and that loudness is not additive, it may be true but it adds nothing to the facts. If it is not consistent, it is simply false. Theories may explain facts; they cannot explain them away. If it should ever prove possible to count the afferent impulses, the theory might be converted into a new set of facts. But we cannot see how they could help. If the relation of the number of afferent impulses excited by a sound X to that excited by a sound Y were the same for all observers, the number may be a function of intensity, but it cannot possibly be a function of directly estimated loudness. If it were not the same for all observers, how should we be advanced in our search for relations valid for all observers?

(9) (Mr Cosens) No questions could fail to be leading questions, in our sense of the words, if the word "twice" were suggested at all to the observer. (Mr Davies) There is no blunder "in assigning the numeral 4 to C," unless at the same time "relation" is identified with arithmetical difference. If relation means arithmetical ratio, of course the assignment is right.

(10) We regard the physical errors of Fechnerism as even more disastrous than the mathematical; that is to say, the belief that all equations can be physically integrated, or that non-additive magnitudes can be added.

(11) We repeat that we do not regard a scale as necessarily useless because it is not based on facts. We accept fully Mr West's and Dr Rein's contention that certain scales are very useful in co-ordinating certain limited groups of observations. But we think they will find that their convenience arises merely from their providing a short name for a relatively complicated function of intensity; and that the observations could be described completely, though more cumbrously, without reference to anything but intensity.

(12) Finally we must deal with Mr Aldridge's plea. No scale would be satisfactory, in the very limited sense of this paper, unless it were wholly based on facts. If the facts on which a satisfactory scale could be based are not at present available, we could do what he asks us only by providing new facts. New facts can be provided only by discovery or by creation; for the present we are leaving these functions to others more competent than ourselves.

THE RECTIFYING PEAK VOLTMETER AS A STANDARD INSTRUMENT

By A. T. Starr, M.A., B.Sc., Faraday House, London

Addendum to paper previously published; received September 9, 1934.*

In § 6 of the paper as published the error due to the capacity of the screened leads from the standard condenser to the rectifier was discussed. The lead-capacity was shown to have a shunting effect whose value was calculated: it has, however, a very important effect on the error due to overbiasing, and discussion of this was unfortunately omitted.

Figure 13 of the paper shows the actual configuration of the rectifying peak voltmeter and figure 14 shows an exact equivalent. Appendix I shows that the overbiasing produces a constant error of e volts in the latter configuration, so that the error is

$$\left(e \div E\frac{C}{C+C_1}\right) \times 100 \text{ per cent,}$$

i.e.

$$\left(e\frac{C+C_1}{C} \div E\right) \times 100 \text{ per cent.}$$

The effect of the lead-capacity C_1 is thus to increase the overbias error from e to $e(C+C_1)/C$.

Dr L. G. Brazier discovered this effect experimentally. He used a sphere gap of 10 $\mu\mu$F. as condenser C, and the lead-capacity C_1 was 5000 $\mu\mu$F. The rectifying system was full-wave, the valves were LP. 4, the grid bias was 9 V. on each valve, and the overbias was in the neighbourhood of 70 V. Dr Brazier found that there was a constant error of 25 kV. r.m.s. with this arrangement. The theory given above shows that the overbias of 70 V. is multiplied by 501 because of the lead-capacity, so that the error is as follows:

$$\begin{aligned} 70 \times 501 &= 35 \text{ kV. (peak)} \\ &= 35/\sqrt{2} \text{ kV. r.m.s.} \\ &= 25 \text{ kV. r.m.s.} \end{aligned}$$

This agrees very well with the experimental values. It was in fact the experiment which led to the discovery of the omission in the theory, and the author wishes to thank Dr Brazier for having thus directed him to the omission.

The fact that the error due to overbiasing was constant at 25 kV. shows that the curvature of the valve characteristic had no appreciable effect.

In conclusion we can say that the lead-capacity may be as high as 20,000 $\mu\mu$F. so far as the shunting of the rectifier is concerned, but it must not be so large that $e(C+C_1)/C$ is an appreciable error. In practice the bias should be adjusted to be a little more than that required to suppress the zero current†. Then the bias should be intentionally increased by, say, 50 per cent or 75 per cent‡. If no change occurs in the rectified current, the error due to overbiasing is negligible.

* *Proc. phys. Soc.* **46**, 35 (1934).

† With LP. 4 valves a grid bias of 4·5 V. is required.

‡ With LP. 4 valves 1·5 or 3 V.

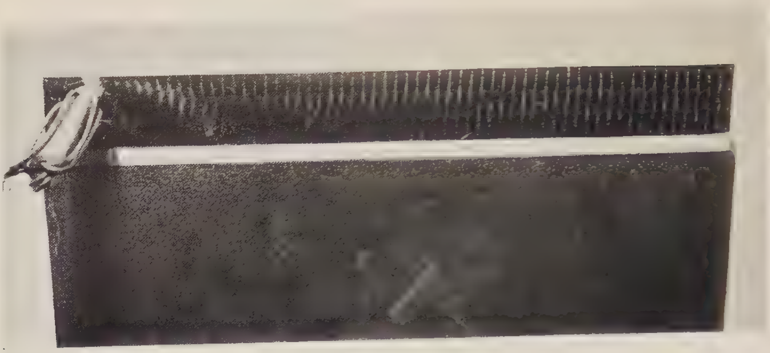


Figure A.

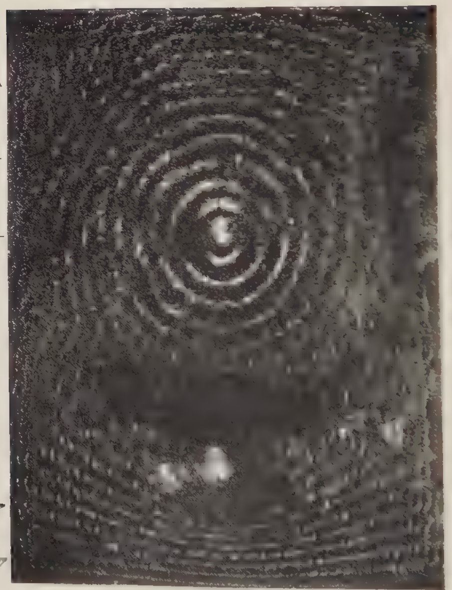


Figure B.

DEMONSTRATIONS

"The velocity of sound in sheet materials." *Demonstration given by* A. B. Wood, D.Sc., F.Inst.P. and F. D. Smith, D.Sc., A.M.I.E.E. *November* 16, 1934.

A recent paper* on this subject describes a method of setting sheets of various materials into transverse vibration by bringing a vibrating nickel tube into contact with them. The reflection of these waves from the edges of the sheet produces a pattern of stationary waves which is made visible by sand scattered on the plate, after the manner of Chladni's figures. The separation of the nodes in the pattern is of the same order as the half-wave-length for transverse vibrations. The high frequency of vibration used by the authors results in a small wave-length and a pattern with a fine structure: in effect, a sheet of moderate size is large in comparison with the wave-length. The fact that the extent of the sheet includes a large number of wave-lengths has practical advantages. If the transverse vibration is subject to appreciable attenuation, as with rubber, cardboard, paper, ebonite and similar materials, the waves reflected from the remote edges of the sheet are practically extinguished: those reflected from the near edges produce simple interference patterns from which numerical values of the velocity of sound can readily be calculated.

To demonstrate the results described in the paper, sand patterns were formed on sheets of various materials by bringing a magnetostriction oscillator vibrating at 20,000 c./sec. into contact with the sheets. Complicated patterns of the usual Chladni type were formed on resonant materials, figure 1 of the paper: and rectangular hyperbolas were obtained on ebonite, figure 3 of the paper. The hyperbolas can be used to calculate the velocity of sound in the material.

In the demonstration a method of obtaining a simple pattern in a resonant material such as aluminium was shown, figure A. The aluminium sheet, of thickness 0·031 in., was damped by placing it between two sheets of spongy rubber except for a narrow strip along one edge. A light and uniform pressure was maintained on the damped portion of the sheet by a slab of solid rubber resting on the top. The undamped exposed strip was set into transverse vibration by bringing a magnetostriction oscillator into contact with it. The suppression of waves within the damped part of the sheet, without reflection at the damping edge, results in a simple pattern of parallel lines within the undamped part. Since these are spaced at half-wave-length intervals, the velocity of sound in aluminium can readily be found.

If the magnetostriction oscillator is applied to the sheet at an angle instead of normally, more complicated modes of vibration involving tangential movement of the sheet are excited. A curious pattern formed in this way on a sheet of keramot†, 0·031 in. thick, was shown, figure *B*. This material has appreciable internal damping and it is certain that the pattern is not formed by interference with edge-reflected

* Page 149 of this volume.

† A commercial insulating material.

waves. Similar patterns are also formed in other sheets, if edge-reflected waves are prevented by internal or external damping. The wave-length in the pattern is different from that of the simple transverse waves, as shown by the hyperbolas in the same photograph. It seems that the pattern is due to interference between vibrations of two types propagated outwards with different velocities from the point of excitation. One of these is probably a transverse vibration of the type considered in the paper and is excited by a force normal to the surface. The other is excited by a force tangential to the surface. The pattern is produced only when both components are present. Provided that the magnetostriction oscillator is applied normally to the sheet so as to excite simple transverse waves only, the anomalous patterns do not occur.

In figure *B* the oscillator was placed in contact with the point marked *O* and inclined at an angle of about 45° to the line *SO* and at right angles to the line *EOW*. The normal component of force excited a transverse vibration radiating uniformly from *O*. The tangential component of force excited a tangential vibration radiating mainly in the directions *ON* and *OS*; these vibrations were in opposite phase, with the result that the interference patterns on the opposite sides of the line *EOW* were complementary, the nodes of one corresponding to the antinodes of the other. As there was no tangential force in the directions *OE* and *OW*, there were no interference patterns in their vicinity.

The transmission of high-frequency vibrations through a long thin wire mechanically connected to the magnetostriction oscillator was also demonstrated, the remote end of the wire producing transverse vibrations in sheets as in the cases mentioned above.

538.213

"A simple apparatus for measuring the pull between two magnetized surfaces."
Demonstration given by Prof. J. T. MacGregor-Morris and C. R. Stoner, B.Sc. (Eng.), A.M.I.E.E. *October* 19, 1934.

Introductory. A comprehensive examination of electrical-engineering and physics laboratories reveals the fact that experiments illustrating the relation between the magnetic flux and the pull between magnetized surfaces are seldom performed. The present investigation was started many years ago in the hope of meeting this need. The aim which the authors have had before them was to produce a robust and at the same time a fairly accurate piece of apparatus. A few laboratories have magnetic-pull permeammeters, but in some of these the apparatus is not used owing to its clumsiness in operation; and generally where any determined attempt has been made to correlate the pull with the flux, erratic results have been obtained and the value of the pull has been found to be too high, especially at low flux-densities. The apparatus about to be described was no exception as far as these matters were concerned, until a prolonged and concentrated effort had been made to master the difficulties.

The pull between two magnetized surfaces as related to the flux passing across the gap has been examined experimentally by several workers. The theoretical re-

lationship was proved by Maxwell to be $P = B^2A/8\pi$, where P is the pull in dynes, B the flux-density in lines per cm^2, and A the cross-sectional area of the gap. Early experimenters obtained erratic results which were always higher than the values calculated from the formula, especially at low flux-densities. Later Threlfall discovered that this was due to the surfaces tending to tilt when parting; this increased the flux-density over a part of the surface, and although the area was decreased, yet because $P \propto B^2$ the pull might be greater under such conditions than with a lower B over a larger area. This effect is found to be particularly noticeable when B is small, because then the reduction of area brought about by the tilt does not produce saturation. Taylor-Jones, using guides and other refinements, obtained consistent experimental results which differed by only $\frac{1}{2}$ per cent from the calculated values between $B = 6000$ and $B = 10{,}000$.

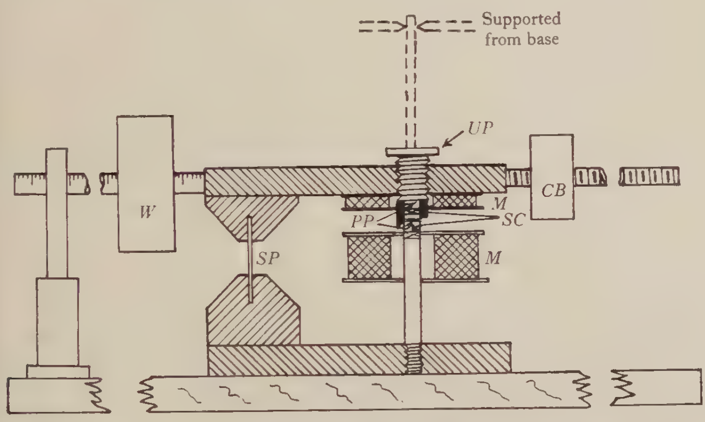


Figure 1. Magnetic-pull apparatus. M, M, magnetizing coils; SC, search coils; PP, surfaces which part; UP, upper pole-piece; SP, spring; W, weight; CB, counterpoise; D, non-magnetic disc; C, collar. [The left-hand drawing has been reduced to $\frac{1}{4}$, but the right-hand drawing has not been reduced.]

A number of workers have interested themselves in the problem of producing a permeammeter in which the value of B should be measured by the pull produced. Although, as we have seen, it is possible to produce such experimental conditions that agreement with theory is obtained, yet it is difficult to embody these conditions in a robust permeammeter. The important problem of flux-leakage has been almost entirely neglected in the permeammeters that have been developed, and either all the ampere-turns are assumed to be expended upon the specimen, or resort is made to a rather unsatisfactory calibration.

Description of the authors' apparatus. For their experimental investigations the authors have constructed the apparatus illustrated in figure 1. This has been designed to permit of comparisons between values of magnetic flux obtained from pull and search-coil measurements, and also an investigation has been begun into its possibilities as a permeammeter. Only the comparison work however will be discussed here.

It will be seen that the magnetic circuit is completed through a stalloy spring which is treated as a fulcrum; the dimensions are $15 \times 1{\cdot}9 \times 0{\cdot}06$ cm. The effect of

this spring when the apparatus is in use can be examined in figure 2, which shows the manner in which the pull between the surfaces falls off as the separation increases. This has been obtained from tests with different thicknesses of non-magnetic separators as described later. It will be seen that when the surfaces part, the decrease in attraction is so much more rapid than the decrease in the effective moment of the weight that the mechanical stiffness of the spring can be neglected.

The top coil is mounted upon the top beam and therefore any attraction between the current-carrying coils is added to the magnetic attraction. This force was both

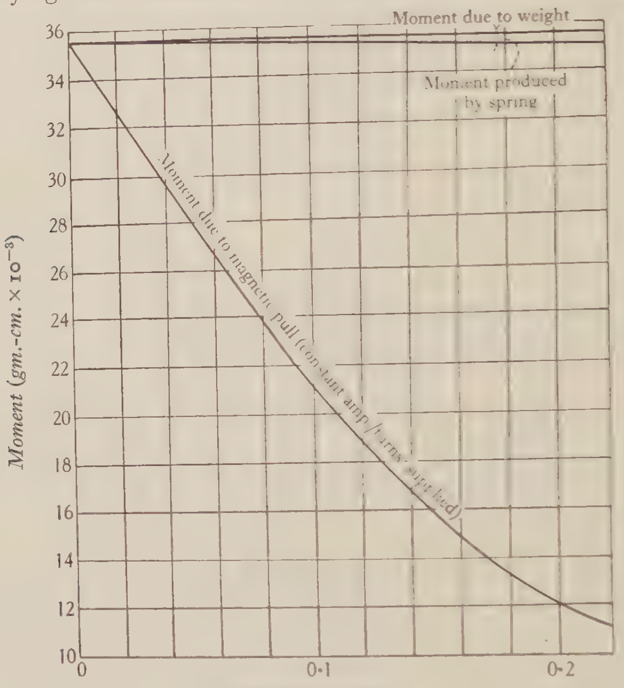


Figure 2.

calculated and measured very approximately and found to be exceedingly small—less than the constraint produced by the spring. The attraction of the top beam by the lower coil was also found by experiment to be negligible.

Method of measuring flux-density and pull. Two similar search coils having an equal number of turns are wound one on each pole-piece. These are first opposed in a ballistic galvanometer circuit and the currents in the magnetizing coils are adjusted till minimum deflection of the galvanometer is obtained on reversal, this being less than 1 per cent of the deflection produced by either search coil alone. The search coil on the lower pole-piece is then used alone to measure the flux-linkages and hence B is obtained, the galvanometer being calibrated each time by means of an air-cored toroid of known dimensions.

The pull is measured by cautiously moving the weight along the beam until the magnetized surfaces part, noting the distance at which this occurs, and hence, by

calculating moments, finding the pull at the gap. The current is then switched off, the gap is closed by sliding back the weight, and then the current is switched on in the reverse direction. In this way the magnetic circuit is kept in a cyclic state and readings can be repeated.

Results. In the apparatus in its original form comparison between B as measured by pull and the same quantity as measured by ballistic galvanometer showed erratic results, the pull-measurement being always too high; errors as great as 20 per cent were common at $B=6000$, and 10 per cent at $B=12,000$. Successive readings of the pull were as a rule consistent among themselves, but readings taken at different times usually showed quite different percentage errors. A systematic check of

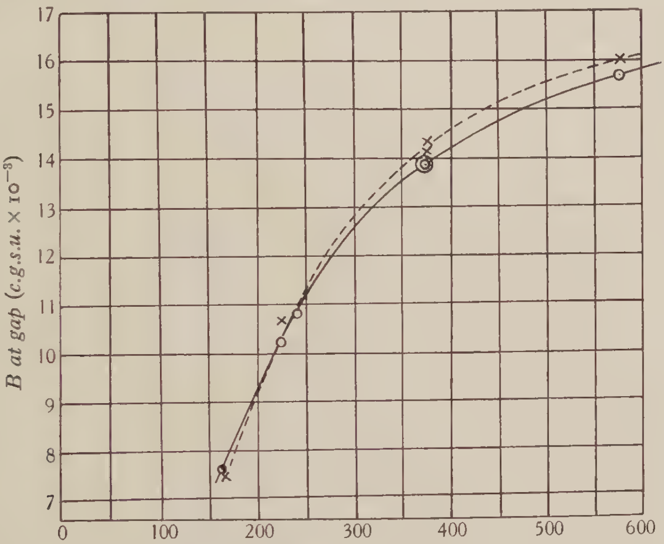


Figure 3. A typical set of results.

possible errors in the galvanometer-calibration etc. was made and failed to reveal any discrepancies.

A study of the work of Threlfall and Taylor-Jones suggested that the irregularities were due to tilting on parting, and so a collar with three guiding claws was devised. A rod was also screwed into the top of the upper pole-piece and this slid through a ring-shaped guide supported from the base of the instrument as indicated by the dotted lines in figure 1. The screw thread on the upper pole-piece was opened out so that the guides could maintain a vertical movement of the pole-piece over a sufficient distance. The apparatus was now assembled while a magnetizing current was flowing. All constraints having been removed, the pole-piece was allowed to bed itself down magnetically into its correct position, and the guide was then locked in its place. A light was placed behind the gap so that the pole-piece and specimen could be seen to be in the correct position. The results obtained were disappointing, for they were as variable as before and errors were of the same magnitude.

It was then decided to insert a thin disc of non-magnetic material between the

pole faces, so as to produce a small permanent gap with the object of reducing the cause for tilting. Now the errors were found to be greatly diminished, as figure 3 shows. Discs of various thicknesses were tried and figure 4 gives the percentage error for a number of discs of different thicknesses operating at various flux-densities. It will be seen that the use of a disc between 0·1 and 0·2 mm. thick brings the agreement between pull and ballistic-galvanometer measurements to within 5 per cent for values of B above 6000. The use of a thicker disc makes the pull-measurement deficient, owing in all probability to magnetic fringing.

Tests were then made with a 0·11 mm. disc, and on removal of the collar and guide the results were found to be almost identical with those obtained when the

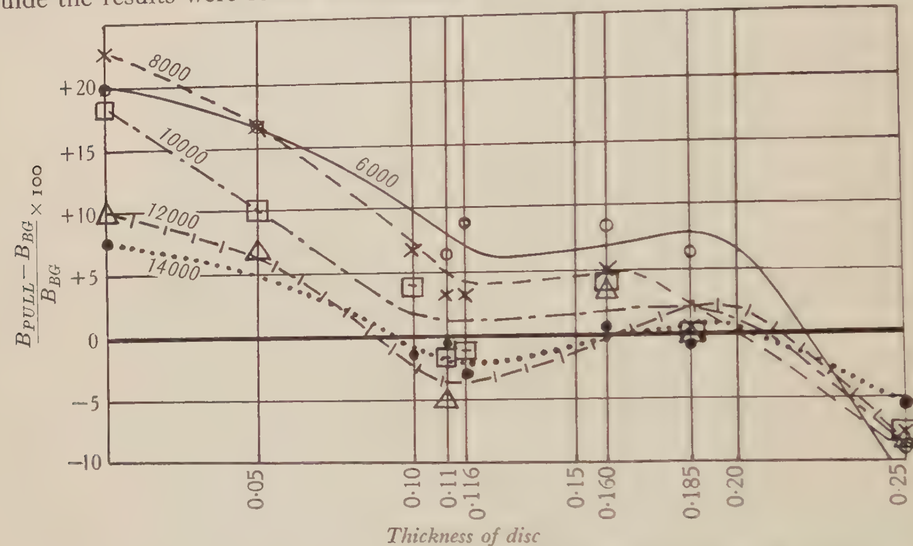


Figure 4. Curves of percentage error for different thicknesses of discs (in mm.).

collar and guide were in use; it proved however desirable to retain the collar, as otherwise the disc occasionally worked out of the gap during the progress of a test.

These results appear to prove that the guide was insufficiently rigid to prevent the extremely small tilt which caused the error, and the insertion of a non-magnetic disc provides a much simpler solution.

It may be objected that it is bad policy to insert a definite air-gap in the magnetic circuit, but it should be remembered that no matter how perfectly the surfaces are worked up from the magnetic point of view there is always an equivalent air-gap, of small dimensions it is true, but still quite measurable in its magnetic effect and varying with the flux-density.

It appeared to be better, therefore, to insert a definite gap intentionally so that its effect would be more nearly constant at all flux-densities and at the same time the errors due to incorrect parting would be reduced to a negligible amount.

Conclusions. The apparatus has now been in the hands of students for some time and consistent results are obtained. It is believed to be more convenient and more accurate than any other apparatus which the authors have seen for the demonstration of the tractive force between magnetized surfaces.

REVIEWS OF BOOKS

Higher Mathematics for Engineers and Physicists, by I. S. SOKOLNIKOFF and E. S. SOKOLNIKOFF. Pp. xiii+482. (New York and London: McGraw-Hill.) 24*s*. net.

We know what to expect in a book entitled "Mathematics for Engineers"—a discussion of how many figures to retain in multiplying 19·34 by 14·17, a long disquisition on the differential equation $\ddot{y}+k^2\dot{y}+m^2y=\cos nx$, a bit about imaginaries, and all the rest of it. This is not one of the family. It is a book on *Higher* Mathematics for Engineers, and is an admirable introduction to its subject. For some reason (or rather, without reason), the authors apologize for addressing the book to physicists as well. Possibly conditions are different on the other side of the Atlantic, but certain it is that if every British physicist mastered this book, few of them would have wasted their time.

The authors plunge into their subject with a chapter on elliptic integrals, and then deal with the numerical solution of algebraic and transcendental equations. Horner's method is included, but one regrets to notice that the method of iteration is not mentioned. After a somewhat sketchy introduction to determinants and matrices, from which the multiplication law for matrices is omitted, we come to one of the best chapters in the book. This is on infinite series, and gives most of the common tests of convergence, deals clearly with uniform convergence, and even gives an example where Taylor's development of a function does not sum to the generating function. Then come chapters on partial differentiation, Fourier series, multiple integrals (with Jacobian introduced), line integrals and improper integrals, and ordinary differential equations. The chapter on the latter again is noteworthy for its thoroughness, dealing as it does with non-linear equations of the second order, and with the method of variation of parameters. It is useful to find in the chapter on line integrals a proper investigation of the conditions under which the value of the integral is independent of the path.

The chapters on partial differential equations, vector analysis, and probability and curve-fitting are less distinguished, whilst the final chapter, on conformal representation, is again very good. This chapter is extracted from a lecture by another author, but one feels after reading the rest of the book that the Sokolnikoffs could have written one at least as good, and the rather noticeable change of style would not then have occurred.

The book is well stocked with examples, mostly rather easy, and the publishers are to be commended on their foresight in printing the words "first edition" on the title-page.

J.H.A.

Mathematical Problems of Radiative Equilibrium, by EBERHARD HOPF. Pp. vi+105. (Cambridge Tracts in Mathematics and Mathematical Physics, No. 31. Camb. Univ. Press.) 6*s*. net.

This tract treats of problems connected with a certain integral equation which occurs on the theory of the outer layers of a star. To quote the preface: "The astrophysicists mostly contented themselves with an approximate solution of these problems. Owing to a certain inherent beauty, however, they aroused also the interest of the rigorous mathematician. It is the purpose of this tract to attempt a coherent representation of all that has been achieved in the direction of a rigorous solution of those standard problems." For the purpose in view, the tract is of high excellence.

In reading it, I have to struggle with a feeling that it is rather sacrilegious to treat a star in this way. Of course, it is not an actual star that is butchered to make a pure mathe-

matician's holiday; it is a "gray" star invented by Schwarzschild, or a purely scattering star invented by Schuster. I fear the reader may not realize that, although the astrophysicist does not treat these problems rigorously, he treats them more rigorously than an actual star is likely to do. Certainly the rigorous mathematicians have secured one triumph. There had been a kind of competition to give the best approximation to the ratio of the boundary temperature to the effective temperature in a gray star. The fourth power of this ratio was first given by Schwarzschild as 0·5; there were four rival later approximations ranging between 0·42 and 0·44. Hopf and Bronstein, by difficult mathematics which occupies many pages in this tract, found that the exact value was $\sqrt{3}/4$.

It may not be out of place to recall that Sir Arthur Schuster, whom we have lately lost, was the pioneer on the astrophysical side of this subject, and his paper "Radiation through a Foggy Atmosphere" (1905) is continually referred to by modern writers as a classic.

A.S.E.

Theoretical Physics, by G. Joos. Translated from the first German edition by I. M. Freeman. Pp. xxiv + 748. (London: Blackie and Son, Ltd.) Price 25*s*.

The first German edition of this book, published in 1932, had an immediate success, and the second edition is already announced. A new chapter on nuclear physics, which is being added to the second German edition, is incorporated in the English translation.

The scope and general nature of the book are already well known to many English users, and the briefest description may suffice here. The text opens with a seventy-page mathematical introduction. This is followed by six sections dealing in turn with mechanics (including relativity mechanics), the field theory of electro-magnetic and optical phenomena, the atomistic nature of electrical phenomena, the phenomenological part of heat theory, the statistical part of heat theory (including quantum statistics), and finally nuclear, atomic and molecular structure and spectral theory.

There are, of course, omissions, but the material has been very well selected, and the treatment is admirably suited to the needs of the average serious student of physics. Here and there the text is supplemented by examples, to which worked solutions are supplied. It is not to be expected that a book of this size should give an adequate knowledge of advanced physics, but Prof. Joos has written a text which may serve as an excellent introduction to more detailed study or as a valuable synopsis of knowledge properly acquired in a less concentrated form.

The publishers have hit upon the singularly happy idea of employing a translator with a knowledge of the subject *and* of the two necessary languages; in brief, the translation is good, and we only rarely encounter passages which might have been clearer if the translator had been willing to take greater liberties with the original phrasing. The book is well produced, and is remarkably good value at the price.

The one adverse criticism, which applies to any sound translation of a good *Lehrbuch*, is that the translation plays too much into the hands of those science students who are already only too content to remain bi-, uni-, or even semi-lingual. This is a pity.

H.R.R.

Atomic Structure and Spectral Lines, Vol. I, by A. Sommerfeld. Translated from the fifth German edition by H. L. Brose, M.A., D.Phil., D.Sc. Pp. xi + 675 with 151 diagrams. (Third English edition. London: Methuen and Co. Ltd.) 35*s*. net.

In the latest German edition of Sommerfeld's classic work the subject matter is divided into two volumes, volume I, which appeared early in 1932, being a revision of the fourth (single-volume) German edition of the original *Atombau und Spektrallinien*, and volume II

an elaboration of the supplementary work *Wellenmechanische Ergänzungsband* of 1928. Accordingly the new English edition is also to appear in two volumes, the first of which is now before us. Compared with the two previous English editions, this volume is in part an abbreviation and in part an extension; its scope will therefore be pretty well known to many readers, and need not be outlined here. Since, as Prof. Sommerfeld himself remarks in his preface, it is possible to understand the new quantum theory only by building it up from the old, the present volume deals both with the fundamental experimental material and also with the conception of orbits as a means of introducing the quantum numbers and as models for subsequent wave-mechanical calculations. The results established by the older theory are given throughout in the forms appropriate to the newer mechanics, the later proofs being reserved for the next volume. The methods of Hamiltonian mechanics, which previously appeared in the Mathematical Appendix, have now been included in the text, whilst certain other matter has been transferred from the text to the appendix. The treatment of multiplet structures has been revised and extended by the inclusion of Pauli's Principle and electron spin. In view of the extent and importance of the recent developments of the study of band spectra, the treatment of this subject in the forty-page chapter to which Prof. A. Kratzer and Dr K. Bechert have contributed seems rather inadequate. Prof. Sommerfeld, it is true, points out in his preface that this short chapter makes no claim to be in any way complete, and contains only what is essential for wave-mechanics.

It is pleasant to see that the figures in the new edition are supplied with captions which were entirely lacking in the earlier editions. Prof. Brose states in his preface that the author wished the translation not to be too literal, slight modifications being left to the translator's discretion. Comparison of certain passages in the first edition with corresponding ones in the new shows that his exercise of this privilege has improved the work. Occasionally one wishes he had gone even farther in this direction, and abandoned some symbols, words, and phrases with which the German reader may be at home but the English reader is not. We are not, for instance, in the habit of using the words "partial band," "R-band," "P-band," "edge" and "disturbance" (Chapter IX) for "band," "R-branch," "P-branch," "head" and "perturbation" respectively; and as a symbol for moment of inertia we prefer I to J in a text in which the latter is also used for a quantum number. And why retain both j and J for one and the same quantum number in the same chapter? We could also wish that it were permissible for a translator to jettison the misnomer "Bergmann series" (p. 352), since the term symbol mb of former editions is now replaced by nF; even if the late Dr Hicks's designation "fundamental series" had not been adopted, there would have been little difficulty and much justice in coupling another investigator's name with these series, for they were discovered by Saunders and by A. Fowler before Bergmann, working in Paschen's laboratory, investigated them, and their significance was first clearly pointed out by Runge.

This English edition contains two sections which are not in the German editions but have deservedly received the approval of Prof. Sommerfeld. One of them is an excellent six-page account of hyper-fine structure and nuclear moments by Dr E. Gwynne Jones; it is well written and admirably suited to the main scheme of the work. The other is a very welcome addendum of five pages by Prof. Brose describing briefly the advances which have been made in atomic physics in the interval between the appearances of the last German edition and the present volume; the subjects of this addendum include investigations of cosmic radiation and nuclear transformations, and the discoveries of the positron, the neutron and the isotopic constitution of hydrogen.

In so large a volume minor slips, especially in the matter of names and initials, are almost unavoidable; only a few so far have been noticed by the reviewer. "A. A. Millikan" (p. 15), "Giaque" (p. 142), "J. W. Aston" (p. 171), "square brackets" (p. 551), "nucleii" (p. 551), "Ann Arbor University" (p. 562), "Lewin" (pp. 566–7), "H.COOH" (p. 586), and "Birheland" (p. 598). The second sentence of the last paragraph on p. 559

is meaningless, owing, as one finds on looking up the corresponding sentence in the first edition, to the omission of several words. On p. 561 "equidistant sequences of lines" should be "sequences of equidistant lines."

English and American students of atomic physics will feel more indebted than ever to Prof. Brose, and will welcome this new edition even more cordially than the old.

W.J.

The Kinetic Theory of Gases (*Some Modern Aspects*), by Prof. M. KNUDSEN. Pp. 64. (London: Methuen and Co., Ltd.) Price 2*s.* 6*d.*

This book practically reproduces the lectures given by Prof. Knudsen at King's College, London, in 1933. It deals almost exclusively with the author's well-known investigations of phenomena occurring at low gas pressures. As a clear and concise first-hand account of these important researches it is assured of a good reception. The beginner should, however, be warned that "Some Modern Aspects" is the operative phrase in the description of the text—he will find no summary of general kinetic theory. H.R.R.

Geometrische Elektronenoptik, by E. BRÜCHE and O. SCHERZER. (Berlin: Springer.) RM 26; cloth-bound RM 28.40.

Although the general phenomena of the effects of electric and magnetic fields on electric discharges have been known for many years, it was only in 1926 that H. Bush showed theoretically that if such fields possessed an axial symmetry they could be used to focus the paths of electrons derived from a single point of origin. The full significance of this work was not at once realized and, broadly speaking, it was not followed up for some four years; but in the period from 1930 to the present time a great deal of work has been done which is ably summarized in the book under review. Not only have the practical possibilities been studied in the period mentioned, but the realization of the wave-directed characteristics of electrons in motion has enabled Glaser and others to apply Hamiltonian methods to the discussion of the theory.

On the practical side, the new methods have proved of the greatest value for the study of thermionic emission from surfaces, and have had valuable applications in the design of cathode-ray oscillograph tubes. It has also proved possible to obtain real images of objects irradiated by electron streams; these will bear a magnification of 8000 or more, even with the employment of magnetic fields for which no attempt has been made to eliminate the aberrations which are known to be present. Theory indicates an ultimately possible resolving-power far out-reaching that of optical microscopy, and it is by no means impossible to deal with biological objects by special methods. The book presents a very complete picture of such practical developments up to the present date, and gives useful chapters on the theory, although the theoretical presentation makes a less coherent whole. The wave-mechanical theory, in spite of its mathematical beauty, proves over-strenuous, and more progress can be made for practical purposes by more straightforward methods.

Since there is such a complete analogy with Gaussian Optics in many respects, we may, perhaps, ask in passing why more attention is not given to the symmetrical form of the paraxial equations. A refractive index appears, for example, as proportional to the square root of a potential, and it would seem to be generally useful to quote the conjugate relation in such a form as

$$\frac{\sqrt{\varphi'}}{l'} - \frac{\sqrt{\varphi}}{l} = \frac{1}{8}\int_{-\infty}^{+\infty} \varphi_z^{-\frac{3}{2}} (\partial\varphi_z/dz)^2\, dz = \frac{\sqrt{\varphi'}}{f'},$$

where l and l' are conjugate distances. The corresponding Gaussian equation is

$$\frac{n'}{l'} - \frac{n}{l} = \frac{n'}{f'} = -\frac{n}{f},$$

if a well-known sign convention is used. Such equations convey far more information than those usual in the book, which merely give an expression of the "focal length" as the result of a theoretical discussion.

In the discussion of such a large range of theoretical and practical matters a good deal is bound to lie outside the practical experience of any author, and it will not be surprising if such a book presents, on the whole, too easy an appearance to the whole subject. The phenomena of the discharge tube are not by any means completely mastered, and although what appear to be results of surprising promise in "microscopy" have already been achieved, there can be few subjects in which real advances will require more patience, skill, and technical resource. Consider some of the factors involved in the present image-production; the very small angular apertures of the beams which have so far been found practicable for use, the heterogeneous character of the electron-energies, the aberrations of the lenses, only measurable with difficulty, and the various difficulties of vacuum work where high vacua must be maintained. It will be seen that the superficial reading of a text-book may easily produce a mental picture glowing in over-rosy hues. On the other hand the book does present a picture of a technical method which, although difficult, reveals no such unsurmountable barrier as that produced by the wave-length of light in optical microscopy; no ultimate barrier, that is, in dealing with smaller and smaller objects until the region of the phantom electron wave-length is reached, in the neighbourhood of 10^{-8} cm. or less. If technical methods of dealing with biological objects are successfully developed, the possibilities of advance are incalculable, and methods of microscopical examination of bodies far beyond resolving-power even of the ultra-violet microscope seem to be within the range of expectation. L.C.M.

Diffraction of X-rays and electrons by amorphous solids, liquids and gases, by J. T. Randall, M.Sc. Pp. xii + 290. (London: Chapman and Hall, Ltd.) 21*s*. net.

This book appears at an opportune moment. The discussions on the solid state at the recent conference showed that the centre of interest is shifting from a consideration of the structure of regular crystals to that of the irregularities which disturb the lattice and of the various non-crystalline and quasi-crystalline structures with which this book deals. In particular the problem of the transition from a liquid or a glassy state to perfect crystalline arrangement seems likely to occupy an increasing amount of attention in the next few years.

The author has performed a valuable task in collecting together the very large amount of scattered work which bears on these problems and presenting it in a clear and logical order. His work will save a great deal of time and trouble to all working in this field and enable them to see more clearly the directions in which to advance.

Apart from its purely scientific interest, the subject is of considerable industrial interest, especially as regards the study of glasses and organic fibres, so that it is appropriate that the author should be a member of one of the great commercial research institutions. The book is extremely up-to-date, and it is interesting to see the stress laid on electron-diffraction as an experimental method which has already proved its usefulness in this field and seems likely to have a rapid development. On the theoretical side the chief results are given with an indication of the way in which they have been obtained, but as a rule without detailed analysis. G.P.T.

Thermionic Emission, by Arnold L. Reimann. Pp. xi + 324. (London: Chapman and Hall, Ltd.) 21*s*. net.

In a volume of 324 pages the author gives an admirable survey of a subject which has received a vast amount of attention in recent years, particularly in the industrial research laboratories of manufacturers of thermionic valves, and it may be mentioned that the author is on the staff of the G.E.C. laboratories.

In the last decade big additions have been made to our theoretical and practical knowledge of thermionic phenomena. The classical theory has been replaced by the quantum statistics in the theory of electrons in metals and in a vacuum, and by the application of the wave-mechanical theory of the transmission of electrons through potential barriers. On the practical side the employment of modern high-vacuum technique has resulted in a marked increase in the quantity and accuracy of the data available.

The volume is primarily concerned with thermal electron-emission and the principal emitters are the metals and certain other electronic conductors. Thirty-three years have elapsed since O. W. Richardson derived his emission formula to fit the experimental results for the emission of electrons from metals. In his theory the velocity-distribution of the electrons available for emission was assumed to be the Maxwellian. Later this theory was involved in so many difficulties that it had to be abandoned and has been replaced by the theory of electrons in metals originally put forward by Pauli and Sommerfeld and since modified and extended by Bloch and others. This theory assumes the existence of effectively free electrons, whose velocity-distribution, instead of being Maxwellian, is governed by the new quantum statistical principles of Fermi and Dirac.

It has been shown by Nordheim and Sommerfeld that the substitution in Richardson's derivation of his original formula of the Fermi-Dirac distribution for the Maxwellian distribution leads to an emission formula concerning whose essential validity there can be no reasonable doubt. It is interesting to note in this connection that H. A. Wilson deduced an emission formula by thermodynamic reasoning without any assumption as to the actual mechanism of the emission, and this theory was further developed by Wilson and Richardson. Later developments of the thermodynamical theory are due to M. V. Laue, W. Schottky and S. Dushman. In its final form the formula derived thermodynamically agrees perfectly with that obtained by Nordheim and Sommerfeld by the quantum statistical method referred to above.

Hence it will be apparent that the study of thermionics touches directly on the electrical constitution of conductors and that the thermal emission of electrons is related to other electron emission phenomena such as the photoelectric effect. In the present volume the theory is presented in a simple form, and where it has not been possible to discuss in any detail such branches of physics as quantum statistics and the modern electron theory of metals, the author has tried to show by what processes of reasoning the results have been arrived at.

The book should prove of great value to those interested in the various practical applications of thermionic phenomena, and the author is to be warmly congratulated on the results of his labours in the production of the volume under review.

E.G.